

Zur Methodik der Planung
und Bewertung diagnostischer
Tumornachsorgeprogramme

Christian Thieme

*Duss
München
1985*



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554620_0063

Aus dem Institut für
Medizinische Informationsverarbeitung, Statistik und Biomathematik
der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. K. Überla

**Zur Methodik der Planung und Bewertung
diagnostischer Tumornachsorgeprogramme**

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Christian Thieme
aus Niederpöcking
München 1985

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. D. Hölzel

Mitberichterstatter: Prof. Dr. H. Denecke
Prof. Dr. H. Ehrhart

Dekan: Prof. Dr. W. Spann

Tag der mündlichen Prüfung: 23.9.1985

Inhaltsverzeichnis

	Seite
1. Fragestellung	1
2. Bestandsaufnahme: Tumornachsorge	6
2.1 Stellung der Nachsorge in der Behandlung Tumorkranker	6
2.1.1 Grundbegriffe aus der Onkologie	6
2.1.2 Definition und Zielsetzung eines Nachsorgeprogramms	10
2.2 Probleme der Planung von Nachsorgeprogrammen	12
2.2.1 Medizinischer Nutzen durch Nachsorge	12
2.2.2 Mit der Nachsorge verknüpfte Risiken	14
2.2.3 Probleme der Finanzierbarkeit	16
2.3 Probleme der Evaluation von Nachsorgeprogrammen	17
2.3.1 Bewertungskriterien für Nachsorgeprogramme	18
2.3.2 Studienansätze	20
2.4 Analogien zwischen Screening und Tumornachsorge	22
2.5 Derzeitige Situation der Tumornachsorge	24
2.5.1 Tumornachsorge an Kliniken und Tumorzentren	24
2.5.2 Die Nachsorgeprogramme der Kassenärztlichen Vereinigung Bayerns	25
2.5.3 Tumornachsorge im Einzugsbereich des Tumorzentrums München	26
2.5.4 Das Tumorregister München	26
3. Methodische Anforderungen bei der Planung von Nachsorgeprogrammen	28
3.1 Präzision diagnostischer Tests	28
3.1.1 Ziele und Anwendungen diagnostischer Tests	28
3.1.1.1 Diagnostik zur Entscheidungsfindung	28
3.1.1.2 Diagnostik im Screening	30
3.1.1.3 Diagnostik im Rahmen von Verlaufsbeobachtung	31
3.1.1.4 Diagnostik in der Tumornachsorge	32
3.1.2 Güte-Indizes für diagnostische Tests	32
3.1.2.1 Prävalenzunabhängige Indizes	32
3.1.2.2 Prävalenzabhängige Indizes	34
3.1.2.3 ROC-Analyse	36
3.1.2.4 Probleme der Testvalidierung	38

	Seite
3.2	Präzision kombinierter Teststrategien in der Nachsorge 40
3.2.1	Test und Referenzkriterium 40
3.2.2	Parallele Anwendung mehrerer Tests 42
3.2.3	Testwiederholung 45
3.3	Bestimmung der faktisch wirksamen Güteindizes 47
3.3.1	Empirische Ermittlung aus der Routine 47
3.3.1.1	Wahrscheinlichkeit des richtig positiven Befundes 48
3.3.1.2	Wahrscheinlichkeit des richtig negativen Befundes 49
3.3.2	Voraussetzungen und Probleme der Übertragbarkeit externer Schätzer 50
3.3.2.1	Beeinträchtigung von Sensitivität und Spezifität durch Besonderheiten der Zielpopulation 51
3.3.2.2	Fehler bei der Testdurchführung und autonome Eigenschaften des Untersuchers 54
3.3.2.3	Chancen zur Beeinflussung von Sensitivität und Spezifität durch geeignete Organisation des Nachsorgeprogramms 55
3.4	Zur pragmatischen Bedeutung von Sensitivität und Spezifität in der Nachsorge 62
3.4.1	Diagnostische Information als Grundlage der Therapieplanung 63
3.4.2	Diagnostisch krank, therapeutisch gesund? 66
3.5	Prävalenz latenter Tumorprogression in der Nachsorge 69
4.	Ansätze zur Bewertung des medizinischen Nutzens von Nachsorgeprogrammen 78
4.1	Ex-ante-Bewertung ohne empirische Daten 78
4.2	Zielkriterien zur empirischen Evaluation 79
4.3	Randomisierte Studien im Vorfeld der Breitenversorgung 89
4.3.1	Festlegung der zu prüfenden Alternativen 89
4.3.2	Durchführbarkeit der Studie 91
4.3.3	Veränderlichkeit der Randbedingungen 92
4.4	Methodische Ansätze zur Evaluation in der Breitenversorgung 95
4.4.1	Interpretation säkularer Trends 95
4.4.2	Fall-Kontroll-Studien 97
4.5	Modellbildung zur Planung und Bewertung von Nachsorgeprogrammen 102

5.	Anforderungen an die Planung und Bewertung von Nachsorgeprogrammen	104
5.1	Zur Anwendbarkeit der WHO-Kriterien an Screeningprogramme in der Nachsorge	104
5.2	Entwurf einer erweiterten Anforderungsliste	113
5.3	Zur Umsetzbarkeit methodischer Forderungen	118
6.	Modellrechnungen zur Tumornachsorge	121
6.1	Zur zukünftigen Entwicklung der Prävalenz Tumorkranker	121
6.1.1	Extrapolation der Neuerkrankungsraten	121
6.1.1.1	Daten und Annahmen	122
6.1.1.2	Methodik der Extrapolation	123
6.1.1.3	Ergebnisse der Extrapolation	125
6.1.2	Weitere für die Entwicklung der Prävalenz Tumorkranker bestimmende Faktoren	127
6.1.3	Diskussion der Prognosen	130
6.2	Zur aktuellen Prävalenz Tumorkranker in Bayern	132
6.2.1	Daten	133
6.2.2	Methodik der Prävalenzschätzung	134
6.2.3	Ergebnisse der Prävalenzschätzung	136
6.2.4	Folgerungen für die Planung von Nachsorgeprogrammen	139
6.3	Prävalenzschätzung für Metastasen und Rezidive im zeitlichen Verlauf am Beispiel Mammakarzinom	141
6.3.1	Datenlage	141
6.3.1.1	Mammakarzinome im Tumorregister München	141
6.3.1.2	Vergleichskollektiv des Istituto Tumori, Mailand	144
6.3.1.3	Repräsentativität der Daten	145
6.3.2	Organ- und zeitabhängige Metastasierungswahrscheinlichkeiten	146
6.3.2.1	Methodik der Schätzung	146
6.3.2.2	Schätzung der metastasenf freien Zeit	148
6.3.2.3	Schätzung des Auftretens bestimmter Organmetastasen im zeitlichen Verlauf	150
6.3.2.4	Lokoregionäre Rezidive	153

		Seite
6.3.3	Gegenüberstellung der ermittelten Metastasierungsmuster mit einer praktizierten Nachsorgestrategie	155
6.3.3.1	Die Nachsorgestrategie des Tumorzentrums München	155
6.3.3.2	Methoden und Annahmen zur Gegenüberstellung	157
6.3.3.3	Ergebnisse der Gegenüberstellung	159
6.3.4	Diskussion der Schätzung von Metastasierungswahrscheinlichkeiten	167
6.3.4.1	Probleme der Datenlage	167
6.3.4.2	Alternativen zum gewählten Modell	170
6.3.4.3	Aussagekraft der Modellrechnung und mögliche Folgerungen	172
7.	Zusammenfassung	178
8.	Literatur	181

Verzeichnis der Tabellen und Abbildungen

	Seite
A2.1: Modellvorstellungen zum Tumorwachstum	9
T2.1: Fragen zum medizinischen Nutzen eines Nachsorgeprogramms	12
T2.2: Fragen zur Abschätzung der Belastungen und Risiken	15
T2.3: Fallzahlen im Tumorregister München	27
T3.1: Anwendung diagnostischer Tests zur Diagnosestellung	28
T3.2: Anwendung diagnostischer Tests zur Verlaufsbeobachtung	31
A3.1: Konkordanz und Diskordanz von Test und Wirklichkeit	33
T3.3: Prävalenzunabhängige Indizes diagnostischer Tests	34
T3.4: Prävalenzabhängige Indizes diagnostischer Tests	35
A3.2: Prävalenzabhängigkeit des prädiktiven Wertes	36
A3.3: Sensitivität, Spezifität und Schwellenwertfestlegung	37
A3.4: ROC-Kurve zweier Tests	37
A3.5: Positivbereich bei Kombination von Screening- und Referenztest	41
A3.6: Kombination von Screening- und Referenztest	41
A3.7: Kombination von Tests zur Erhöhung der Sensitivität	43
A3.8: Kombination von Tests zur Erhöhung der Spezifität	44
A3.9: Zur Logik der Wiederholung von Untersuchungen	46
T3.5: Einfluß testrelevanter Faktoren in der Zielpopulation auf Sensitivität und Spezifität	51
T3.6: Einfluß von Fehlern bei der Testdurchführung und autonomen Eigenschaften des Untersuchers auf Sensitivität und Spezifität	54
T3.7: Chancen zur Hebung von Sensitivität und Spezifität durch geeignete Organisation des Nachsorgeprogramms	56
A3.10: Überlebenszeit in Abhängigkeit vom Karnofsky-Index	58
A3.11: Wege zur Entdeckung einer Tumorprogression	63
A3.12: Interventionsebenen bei Tumorprogression	65
A3.13: Anwendung zweier Tests mit unterschiedlichen testspezifischen Zielereignissen	67
A3.14: Mögliche Gesamtergebnisse bei simultaner Verwendung von Tests	67
T3.8: Faktoren, die die Prävalenz latenter Tumorprogression beeinflussen	71

	Seite
T4.1: Zielkriterien zur Evaluation des Nutzens systematischer Tumornachsorge	81
A4.1: Karnofsky-Index zur Beschreibung des Allgemeinzustandes	85
A4.2: QL-Index zur Messung der Lebensqualität	86
A4.3: Beispiel zum Wandel des Stellenwertes von Tumornachsorge	93
T4.2: Zur Durchführbarkeit von Fall-Kontroll-Studien	99
T4.3: Mathematische Modelle	102
T5.1: Modifikationen der WHO-Kriterien bei Anwendung auf Nachsorgeprogramme	105
T5.2: Entwurf einer erweiterten Anforderungsliste an Nachsorgeprogramme	114
T5.3: Vorschlag für ein Basis-Nachsorgeprogramm	119
A6.1: Entwicklung des Anteils der über 60-jährigen	124
T6.1: Ermittlung von zur Extrapolation geeigneten Inzidenzen	125
T6.2: Extrapolation der Rate der Neuerkrankungen an malignen Tumoren	126
T6.3: Entwicklung der erwarteten Neuerkrankungen absolut	126
A6.2: Zeitlicher Trend einiger Krebsformen	128
T6.4: 5-Jahres-Überlebensraten zur Schätzung der Prävalenz	134
T6.5: Prävalenz männlich: Prozentangaben vertikal	137
T6.6: Prävalenz weiblich: Prozentangaben vertikal	137
T6.7: Prävalenz beide Geschlechter: Prozentangaben vertikal	137
T6.8: Prävalenz männlich: Prozentangaben horizontal kumuliert	138
T6.9: Prävalenz weiblich: Prozentangaben horizontal kumuliert	138
T6.10: Prävalenz beide Geschlechter: Prozentangaben horizontal kumuliert	138
T6.11: Zusammenfassung: Anzahl Nachsorgekontakte in Bayern pro Jahr	140

	Seite
T6.12: Definition des 'prospektiven Kollektivs'	142
T6.13: Metastasenlokalisation und Fächergruppe des Datenurhebers im retrospektiven Kollektiv	143
T6.14: Metastasenlokalisation prospektiv und retrospektiv	144
T6.15: Altersverteilung bei Diagnosestellung eines Mammakarzinoms für die weibliche Bevölkerung	146
A6.3: Verlauf der Fernmetastasierung	149
T6.16: Schätzungen für λ zur Beschreibung der metastasenfreien Zeit	149
T6.17: Relative Häufigkeiten der einzelnen Metastasen im zeitlichen Verlauf	151
T6.18: Zeitunabhängige relative Häufigkeiten der einzelnen Metastasen	152
T6.19: Muster des synchronen Auftretens von Metastasen	152
A6.4: Postoperative Behandlung des Mammakarzinoms in München	154
A6.5: Die Nachsorgestrategie des Tumorzentrums München	155
A6.6: Ansatz zur a-priori-Evaluation einer Nachsorgestrategie	157
T6.20: Strategiesegmente zur Nachsorge beim Mammakarzinom	158
T6.21: Anzahl der Untersuchungen/Jahr bei landesweiter Anwendung der Nachsorgestrategie des Tumorzentrums München	160
T6.22: Metastasierungsmuster und Nachsorgestrategie: Skelett	161
T6.23: Metastasierungsmuster und Nachsorgestrategie: Lunge/Pleura	162
T6.24: Metastasierungsmuster und Nachsorgestrategie: Mamma	163
T6.25: Metastasierungsmuster und Nachsorgestrategie: Leber	164
T6.26: Alternative zur periodischen Beobachtung der Leber	165
T6.27: Aufwand zur Erfassung isolierten Leberbefalls	166

1. Fragestellung

In den Mortalitätsstatistiken der entwickelten Länder belegen die Krebserkrankungen den zweiten Rang, von dem sie trotz weltweiter Anstrengungen in der Krebsbekämpfung bisher nicht verdrängt werden konnten. Innerhalb der einzelnen Altersklassen hat sich das Risiko, an Krebs zu sterben, in den zurückliegenden Jahrzehnten offenbar kaum verändert (46).

Dies bedeutet, daß die tatsächlich beobachtete Häufigkeit (Crude Rate) der Sterbefälle bzw. Neuerkrankungen sogar zunimmt, wenn sich der Anteil der älteren Menschen in einer Population erhöht. In der bundesdeutschen Bevölkerung wird dieser Trend besonders ausgeprägt beobachtet bzw. prognostiziert.

Eine rasche und umfassende Lösung des Krebsproblems kann nach übereinstimmender Ansicht der in die Krebsforschung involvierten Disziplinen auch in den nächsten Jahren nicht erwartet werden. Daher wird die Prävention der Krebserkrankungen auf allen Stufen des Entwicklungsprozesses auch in Zukunft wissenschaftlich und sozialmedizinisch von Bedeutung sein.

Primäre Prävention beschreibt in diesem Zusammenhang das Bestreben, Karzinogene und Kokarzinogene aus der Umwelt zu entfernen oder Verhaltensweisen abzustellen, die zur Exposition gegen verursachende Faktoren führen.

Unter sekundärer Prävention wird die frühzeitige Erfassung beginnender Krankheitsprozesse in der präklinischen, asymptomatischen Phase verstanden. Voraussetzung zur Entdeckung asymptomatischer Tumorerkrankungen ist die Durchführung von periodischen Reihenuntersuchungen im Rahmen von Früherkennungsprogrammen. In der Bundesrepublik werden seit über 10 Jahren gesetzliche Früherkennungsuntersuchungen für ausgewählte Krebslokalisationen angeboten.

Die Primärbehandlung manifester Tumorerkrankungen wird in diesem Kontext gelegentlich als tertiäre Prävention bezeichnet. Allerdings werden nicht alle Tumorerkrankungen in Stadien diagnostiziert, die noch ein therapeutisches, insbesondere kurativ wirksames Eingreifen ermöglichen.

In formaler Analogie zu sekundärer und vor allem tertiärer Prävention könnten Maßnahmen der Tumornachsorge als quartäre Prävention bezeichnet werden. In welchem Umfang der durch die Bezeichnung Prävention offenbarte Anspruch in der Phase der Tumornachsorge noch realisiert werden kann, ist empirisch noch wenig untersucht. Auch ohne die detaillierte Kenntnis empirischer Daten muß jedoch angenommen werden, daß das Bestreben, den Patienten durch Nachsorge und frühzeitige Rezidivbehandlung zu heilen, höchstens für einen gewissen Teil der Fälle relevant ist. Die Größe dieses Anteils wird nicht zuletzt von der Definition der Nachsorge und ihrer Abgrenzung zur Primärtherapie bestimmt.

Im Mittelpunkt dieser Arbeit steht die im Anschluß an die Primärtherapie routinemäßig und periodisch durchgeführte Diagnostik zur frühzeitigen Entdeckung von Metastasen, Rezidivtumoren, Therapiefolgen und Zweitmalignomen. Unter Nachsorgestrategie soll in diesem Zusammenhang eine Empfehlung über Art und zeitliche Anordnung der durchzuführenden Untersuchungen verstanden werden. Mit Nachsorgeprogramm soll, in Analogie zur Bezeichnung Früherkennungsprogramm, die Anwendung einer Nachsorgestrategie in einem vorgegebenen Umfeld bezeichnet werden. Im Detail wird dies in Abschnitt 2 definiert.

Vor allem in den USA hat sich zur Bezeichnung komplexer Dienstleistungen in der Medizin der Begriff "Medical Technology" eingebürgert, definiert als "drug, device, procedure, equipment, or a combination of these into a system" (11). An der Bewertung solcher Technologien, dem "Medical Technology Assessment", wirken nicht selten zahlreiche Disziplinen mit, wie Volkswirtschaft, Psychologie, Ingenieurwissenschaften und Unternehmensforschung. Zur Planung und Bewertung der "Technologie" Tumornachsorge kann insbesondere auch statistisch-epidemiologische Methodik in vielen Punkten beitragen. Dagegen beinhaltet fehlende oder unzureichend angewandte Methodik das Risiko vermeidbarer Fehlplanungen bzw.

vermeidbar fehlerhafter Schlußfolgerungen. Ziel dieser Arbeit ist es, aus methodologischer Sicht Richtlinien und Anforderungen für die Planung und Bewertung von Nachsorgeprogrammen zu entwickeln.

Die präzisierte Fragestellung lautet

- Bestandsaufnahme der Nachsorge unter methodischen Gesichtspunkten vorzunehmen,
- zu untersuchen, welche methodischen Fehler (Sensitivität, Spezifität, Definition der Zielereignisse, Prävalenz) bei Nachsorgeprogrammen auftreten können und wie sie zu behandeln sind,
- zu untersuchen, inwieweit kontrollierte Studien zur Bewertung von Nachsorgeprogrammen geeignet sind und welche Alternativen ggf. bestehen,
- zu untersuchen, inwieweit die Anforderungen an Nachsorgeprogramme insgesamt durch den WHO-Katalog zur Beurteilung von Screeningprogrammen (71) erfaßt werden und ggf. den Entwurf einer erweiterten Anforderungsliste zu erstellen, sowie
- Modellrechnungen für die Durchführung von Nachsorgeprogrammen in Bayern anzustellen.

Im Rahmen der Bestandsaufnahme in Abschnitt 2 werden die Charakteristika von Nachsorgeprogrammen und deren Stellung in der Betreuung Tumorkranker diskutiert und die in diesem Zusammenhang auftretenden methodischen Probleme präzisiert. Eine Übersicht über laufende Aktivitäten zur Nachsorge Tumorkranker unterstreicht den aktuellen Bezug der anzusprechenden Fragen.

Im Zusammenhang mit der Planung von Nachsorgeprogrammen werden in Abschnitt 3 Anforderungen und Zusammenhänge aufgezeigt, deren Nichtbeachtung Planungsfehler und später verfehlte Schlußfolgerungen nach sich ziehen kann. Wichtige Fehlerquellen sind die unzureichende Auswahl bzw. Interpretation von Planungsgrößen sowie falsche Annahmen über deren quantitative Ausprägung. Der Beitrag der statistischen Methodik betrifft vorrangig die zur Beurteilung von Diagnostikprozeduren benötigten Indizes Sensitivität und Spezifität sowie die Prävalenz unterschiedlicher Zielereignisse.

Untersucht wird einerseits der Stellenwert dieser Größen im Rahmen von Nachsorgeprogrammen sowie deren relative Gewichtung.

Hierzu zählt u.a. die Bereitstellung von Kriterien, wann der Sensitivität der Vorrang vor der Spezifität einzuräumen ist und umgekehrt, oder Überlegungen zur Wahl des adäquaten Zielereignisses als "Nenner" dieser Indizes.

Weiterhin werden Kriterien erarbeitet, wann und mit welchen Einschränkungen extern ermittelte quantitative Schätzer für Sensitivität, Spezifität und Prävalenz zur Planung von Nachsorgeprogrammen heranzuziehen sind und welche Faktoren die Übertragbarkeit solcher Werte auf die in einem Nachsorgeprogramm gegebene Situation beeinträchtigen können. Diese Übertragbarkeit auf ein Nachsorgeprogramm kann teilweise erst in Kenntnis des Umfelds und der organisatorischen Gestaltung des Programms diskutiert werden. Dies wird teilweise dazu führen, daß Anforderungen an die Organisation methodisch begründet werden können.

Für die Bewertung des medizinischen Nutzens von Nachsorgeprogrammen ist von den allgemeinen Grundsätzen zur Durchführung von Wirksamkeitsanalysen auszugehen. Das hierzu aus methodischer Sicht optimale Instrument des kontrollierten klinischen Versuchs zählt zu den intensiv beforschten Zweigen der Biostatistik. Entscheidungsrelevant für die Anwendbarkeit randomisierter Studien in der Nachsorge ist jedoch u.a. die Frage, ob deren ordnungsgemäße Durchführung unter den gegebenen medizinischen und organisatorischen Randbedingungen wahrscheinlich ist.

Für die Bewertung von Screeningprogrammen mit teils vergleichbarer Problematik wird diese Methodendiskussion anhand publizierter Daten geführt. Für Nachsorgeprogramme kann derzeit nur auf Analogieschlüsse und a-priori-Überlegungen zurückgegriffen werden. In diesem Sinne werden in Abschnitt 4 Zielkriterien zur Bewertung des medizinischen Nutzens sowie Chancen und Limitierungen randomisierter Designs und möglicher Alternativen diskutiert.

Ziel des 5. Abschnitts ist der Entwurf einer allgemeinen Anforderungsliste für die Planung und Bewertung von Nachsorgeprogrammen auf der Basis der gewonnenen methodischen Ergebnisse. Hierzu wird von dem bereits in den 60er Jahren von der WHO aufgestellten 10-Punkte-Katalog (71) zur Beurteilung von Screeningprogrammen ausgegangen, der statistisch-methodische, organisatorische, sozialmedizinische und ökonomische Aspekte umfaßt.

In der Diskussion neu zu etablierender Früherkennungsprogramme hat sich dieser Katalog bewährt und wird weitgehend unverändert angewandt. Auch für das enger umschriebene Feld der Tumornachsorge könnten sich solche Kriterien als Orientierungshilfe erweisen. Ein erster Ansatz hierzu, der vorwiegend die statistisch-methodischen Aspekte reflektiert, wird in dieser Arbeit vorgelegt und bezüglich der Realisierungsmöglichkeiten diskutiert.

Angesichts der unzureichenden Datenlage, vor allem zur Prävalenz, wird im 6. Abschnitt abschließend versucht, über Modellrechnungen Wege zur besseren Nutzung vorhandener Datenbestände aufzuzeigen.

Von Interesse sind einmal Daten zum Mengengerüst, um zukünftige Planungen konkreter Nachsorgeprogramme hinsichtlich der erwarteten Inanspruchnahme simulieren zu können. Hierzu wird die weitere Entwicklung der absoluten Neuerkrankungszahlen in der Bundesrepublik bis zum Jahr 2030 extrapoliert. Anschließend wird die derzeitige Prävalenz Tumorkranker in Bayern geschätzt, wobei nach Alter und Zeitraum seit Diagnosestellung geschichtet wird.

Zum anderen erfordert die konkrete Planung einer Nachsorgestrategie Daten zum natürlichen Verlauf der Zielerkrankung, wie sie teilweise aus verlaufsorientierten Tumorregistern gewonnen werden können. Am Beispiel Mammakarzinom wird deshalb versucht, aus Registerdaten Vorhersagen über das zeitliche Auftreten der für die Nachsorge relevanten Folgeereignisse im Zeitraum von 5 Jahren nach Diagnosestellung zu treffen. Die Ergebnisse dieser Simulation werden mit einer aktuellen Nachsorgeempfehlung verglichen.

In einer Zusammenfassung werden die wichtigsten Ergebnisse abschliessend dargestellt.

2. Bestandsaufnahme: Tumornachsorge

2.1 Stellung der Nachsorge in der Behandlung Tumorkranker

2.1.1 Grundbegriffe aus der Onkologie

Vor der klinischen Manifestation einer Tumorerkrankung durchläuft ein maligner Tumor in einem jahre- oder jahrzehntelangen Prozeß verschiedene präklinische Phasen und Stadien. Das Wachstum erfolgt durch unkontrollierte Zellteilung. Ausgehend von einer einzigen Zelle würde ein Tumor die Größe von 1 cm Durchmesser erreichen, nachdem er sich 32mal verdoppelt hat. Nach etwa 40 Verdopplungen ist die biologisch tolerable Tumormasse von 10^{12} Zellen überschritten. Der klinisch beobachtbare Teil des Krankheitsverlaufs entspricht daher in etwa nur dem letzten Fünftel der Entwicklung. Allerdings ist das Grundlagenwissen über Wachstumsverhalten und mögliche Interaktionen zwischen Krankheit und körpereigener Abwehr in der latenten Phase des Tumorstadiums lückenhaft.

Die gefährlichste Eigenschaft eines malignen Tumors ist die Fähigkeit zur Ausbildung von Tumormetastasen. Unter Metastasierung wird die Loslösung von Tumorzellen aus dem ursprünglichen Verband mit nachfolgender Absiedelung in einem räumlich vom Primärtumor getrennten Organ oder Gewebe des Körpers verstanden.

Der Transport dieser Tumorzellen im Körper erfolgt häufig über den Lymphabfluß des primär erkrankten Organs (lymphogene Metastasierung) oder über die Blutbahnen (hämatogene Metastasierung). Als drittes physiologisches Transportmedium für Tumorzellen können Ergüsse, Aszites etc. in Frage kommen. Schließlich ist die nicht 'physiologische' Verschleppung von Tumorzellen zu berücksichtigen, die im Zusammenhang mit diagnostischen oder therapeutischen Eingriffen diskutiert wird.

Ein zusätzliches Kriterium der Bösartigkeit eines Tumors ist die Fähigkeit zum lokal destruierenden Wachstum, das heißt, die Tendenz, vom primär erkrankten Organ aus direkt infiltrierend auf Nachbarstrukturen überzugreifen.

Aus dieser Aufreihung von Eigenschaften allein ist keine genaue Definition von 'Malignität' abzuleiten, denn nicht jeder 'böartige' Zelltyp erfüllt alle diese Kriterien mit der gleichen Wahrscheinlichkeit. Für die überwiegende Zahl als maligne angesehener Tumoren gilt jedoch die Formel:

Malignität = lokale Infiltration + Fähigkeit zur Metastasierung.

Auch die verbreitete TNM-Klassifikation maligner Tumoren (64) ist an dieser Formel festgemacht. Prinzip der TNM-Klassifikation ist die Beschreibung der Ausdehnung einer Tumorerkrankung beim individuellen Patienten anhand der Kriterien

T (Tumor)	Ausmaß der lokalen Ausdehnung und Infiltration
N (Nodes)	Vorliegen von Tumormetastasen in den regionären Lymphknoten
M (Metastasen)	Vorliegen anderweitiger Tumormetastasen (= Fernmetastasen)

Für die interpretierende Bewertung eines TNM-Stadiums ist die Richtung von rechts nach links maßgebend. Wenn bereits Fernmetastasen (M) vorliegen, verlieren regionärer (N) und lokaler (T) Befund an prognostischer Bedeutung. Bei Fehlen von Fernmetastasen wird die Prognose überwiegend durch den Lymphknotenstatus bestimmt. Bei Metastasenfreiheit auch in den regionären Lymphknoten schließlich scheint die Größe des Primärtumors, insbesondere die Einhaltung oder Überschreitung der Organgrenzen, ein guter Gradmesser für das wahrscheinliche Ausmaß latenter Metastasen und damit für die Prognose zu sein.

Für die primäre Therapie maligner Tumoren ist unter Berücksichtigung individueller Gegebenheiten vor allem die erkennbare Tumorausdehnung und die Wahrscheinlichkeit für das Vorhandensein weiterer, latenter Tumormanifestationen bestimmend.

Von kurativer oder potentiell kurativer Therapie wird gesprochen, wenn kein Anhalt für das Vorliegen von manifesten Fernmetastasen besteht, die vorhandene Tumormasse im Rahmen des Erkennbaren (radikal) beseitigt wurde und die Wahrscheinlichkeit für eine latente Generalisierung der Erkrankung 'hinreichend' gering ist. Hierfür sind die Kenntnis des biologischen Verhaltens der betref-

fenden Tumorart im allgemeinen und der lokale Befund maßgeblich. Wenn durch lokale Maßnahmen die Beseitigung des Tumors offensichtlich oder erfahrungsgemäß nicht möglich ist, spricht man von palliativer Therapie. Ziele palliativer Maßnahmen sind Lebensverlängerung, Beseitigung von Beschwerden, Verhütung von vermeidbaren Komplikationen usw.

Eine Bereicherung der Möglichkeiten ergibt sich aus der Kombination lokaler Intervention mit regionären (Bestrahlung) oder systemischen (Chemo-/Hormontherapie) Maßnahmen. Unter adjuvanter Therapie wird die im Anschluß an eine potentiell kurative Operation erfolgende prophylaktische Behandlung von nicht sichtbaren, aber mußmaßlich oder möglicherweise vorliegenden Tumormetastasen (Mikrometastasen) verstanden. Die Nachbestrahlung der Axilla oder die Durchführung einer systemischen Therapie nach chirurgisch radikaler Behandlung eines Mammakarzinoms sind Beispiele für adjuvante Therapiemaßnahmen.

Die Möglichkeit der restlosen Beseitigung aller Tumorzellen durch die Therapie ist derzeit nach Ansicht der meisten Experten nahezu auszuschließen (43). Es muß wohl davon ausgegangen werden, daß der Prozeß der Erkrankung im günstigsten Fall in einen Zustand zurückverlegt wird, in dem ein erneutes Wachstum des Tumors durch die körpereigene Abwehr verzögert oder verhindert werden könnte (Abb. 2.1).

Die Wahrscheinlichkeit, daß die durchgeführte Therapie dennoch in der Lage ist, einen dauerhaften Heilungsprozeß zu initiieren, wird unterschiedlich beurteilt. Nach außen manifestiert sich dies u.a. in unterschiedlicher Terminologie zur Beschreibung des Krankheitsverlaufs nach Primärtherapie. Optimistischere Betrachtung führt zum Postulat einer 'Tumorfreiheit', ggf. gefolgt vom 'Rezidiv', wenn erneut Tumorherde klinisch oder diagnostisch faßbar werden. Bei zurückhaltenderer Beurteilung wird von 'Remission' und nachfolgender 'Progression' gesprochen.

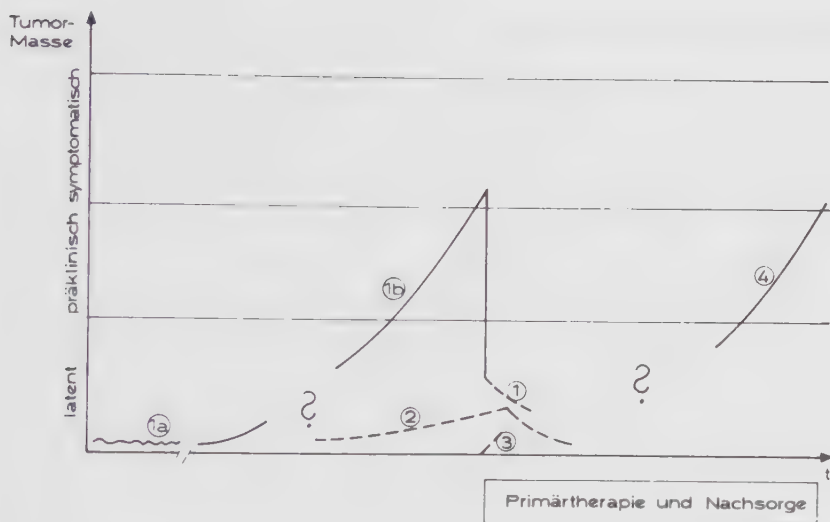


Abb. 2.1: Wachstum eines malignen Tumors.

- 1a: Tumor in der latenten Phase, Auseinandersetzung mit dem Immunsystem
- 1b: Wachsender Primärtumor, Operation
- 2.: Latente Lymphknotenmetastasierung, Nachbestrahlung
- 3.: Intraoperative Dissemination
- 4.: Manifeste Progression

Unabhängig von diesen terminologischen Unterschieden kann der Zustand, in dem klinisch oder diagnostisch keine Tumorzeichen erkennbar sind, durch drei Klassen von Ereignissen oder deren Kombination beendet werden. Es handelt sich um Wiederauftreten des Tumors am Ort der primären Erkrankung (Lokalrezidiv), in den regionären Lymphknoten (regionäres-Rezidiv bzw. Neumanifestation in den regionären Lymphknoten) oder in entfernten Organen und Strukturen (Fernmetastasen).

Diese Ereignisse bilden neben Therapiefolgeerkrankungen und Zweitmalignomen, das sind Tumoren, die nicht als Metastasen des bereits bekannten Primärtumor erklärbar sind, die wichtigsten Zielereignisse für die diagnostische Überwachung im Rahmen von Nachsorgeprogrammen. Bestandteile und Zielsetzung von Tumornachsorgeprogrammen werden im folgenden Abschnitt vorgestellt.

2.1.2 Definition und Zielsetzung eines Nachsorgeprogramms

Gynäkologen waren die ersten, die die Notwendigkeit einer kontinuierlichen Kontrolle operierter Tumorpatientinnen postulierten. Seit Durchführung der ersten Unterleibsoperationen nach WERTHEIM im Jahre 1898 wurden alle nach dieser Methode behandelten Patientinnen jährlich einmal untersucht, so daß im Jahre 1925 über 5-Jahres-Ergebnisse für die ersten 1000 Patientinnen berichtet werden konnte, wobei nur 4 (!) Patientinnen aus dem Follow-up verloren waren (70). Es ist bemerkenswert, daß diese Leistung, erbracht u.a. während des ersten Weltkrieges und der darauf folgenden turbulenten Ereignisse, heute wegen der zunehmenden Mobilität der Bevölkerung - auch bei Unterstützung des Follow-up durch Computer - kaum reproduziert werden kann. Daneben fällt auf, daß schon diese frühe Aktivität gleichwertig von der Sorge um die Patientinnen und dem Bemühen um Gewinnung von Daten über den Stellenwert der Behandlung getragen war. Sicher erzeugte auch die mit 42% anfangs extreme Operationsmortalität besonderen Rechtfertigungsdruck.

Aufgrund erweiterter Interventionsmöglichkeiten sind die Handlungsalternativen und damit auch die Aufgaben für die Tumornachsorge heute vielfältiger. Im weitesten Sinne können unter Tumornachsorge sämtliche ärztlichen und einige paramedizinische Leistungen verstanden werden, die an Tumorpatienten nach Abschluß der Primärtherapie erbracht werden.

Hierzu zählen

- Weiterführung von adjuvanten Therapiemaßnahmen,
- Langzeitbehandlung primär inkurabler Patienten,
- Behandlung von chronischen oder akuten Begleiterkrankungen,
- physische Rehabilitation,
- psychosoziale Betreuung sowie

sowie die in dieser Arbeit im Mittelpunkt stehende

- diagnostische Überwachung.

'Nachsorge und 'Diagnostik in der Nachsorge' werden im folgenden synonym verwendet, wenn aus dem Kontext keine Mißverständnisse zu erwarten sind.

Unter einem Nachsorgeprogramm soll ein Versorgungsangebot an Tumorkranken verstanden werden,

- das sich im Anschluß an die Primärtherapie bis zum Ablauf einer gegebenen Zeit (z.B. 5 Jahre) an klinisch tumorfreie Patienten einer bestimmten Diagnose(-gruppe) richtet;
- das dem Patienten in diesem Zeitraum regelmäßig Kontakt zu seinem Arzt ermöglicht;
- das Art, Abfolge und Frequenz von diagnostischen Tests festlegt, die zur frühzeitigen Erkennung von Ereignissen durchzuführen sind, und über relevante Symptome informiert;
- das Empfehlungen für das Vorgehen bei pathologischen und/oder abklärungsbedürftigen Befunden gibt;
- das Kriterien zur Bewertung der Arztcompliance liefert;
- das Kriterien zur Bewertung des Nutzens sowie zur Fortschreibung und Optimierung liefert;
- das an neue Entwicklungen und Erkenntnisse angepaßt werden kann.

Bestehende Programme entsprechen dieser programmatischen Definition nicht. Insbesondere die Forderungen an Organisation und Lernfähigkeit von Nachsorgeprogrammen beschreiben Zielsetzungen. Die Mehrzahl der publizierten Vorschläge beschränkt sich auf die Definition eines Diagnostikschemas, d.h. Festlegung der anzuwendenden diagnostischen Tests und deren Anordnung im zeitlichen Verlauf zur frühzeitigen Entdeckung von Ereignissen. Ein solches Diagnostikschema wird im folgenden als Nachsorgestrategie bezeichnet. Die Nachsorgestrategie ist somit Kern eines Nachsorgeprogramms, welches gemäß der oben gegebenen Definition jedoch mehr umfaßt als nur ein Diagnostikschema.

Die in einem Nachsorgeprogramm durchgeführten diagnostischen Tests verfolgen vorwiegend drei Ziele, und zwar

- Früherkennung von lokoregionären Rezidiven und Fernmetastasen,
- Früherkennung von Therapiefolgen und
- Früherkennung von Zweitmalignomen.

Die Erwartung, die sich an etwaige positive Befunde knüpft, kann nur tumorspezifisch, häufig nur abhängig von weiteren Detailinformationen formuliert werden. Das Spektrum des möglichen medi-

zinischen Nutzens ist breit; es umfaßt

- Heilung,
- Lebensverlängerung,
- Verhinderung oder Milderung von Komplikationen der Tumorerkrankung,
- Verhinderung oder Milderung von Therapiefolgen,
- somatisch orientierte Verbesserung der Lebensqualität,
- psychologische Stabilisierung sowie
- Gewinnung von Fakten zur Bewertung der Primärtherapie und Optimierung der Nachsorge.

2.2 Probleme der Planung von Nachsorgeprogrammen

2.2.1 Medizinischer Nutzen durch Nachsorge

Neben den erwünschten Effekten beinhaltet Nachsorge auch Belastung und Risiko für den Patienten sowie Kosten für die Allgemeinheit. Bestimmend für die Festlegung eines Programms ist die Abwägung zwischen den unerwünschten Eigenschaften und dem erwarteten Nutzen.

Fragen zum medizinischen Nutzen eines Nachsorgeprogramms sind in der folgenden Tabelle zusammengestellt.

Tab. 2.1: Fragen zum medizinischen Nutzen eines Nachsorgeprogramms

-
1. Welchen Stellenwert hat der auf einen Befund folgende therapeutische Eingriff?
 2. Bringt die frühzeitige Entdeckung durch Nachsorge erkennbare oder mutmaßliche Vorteile im Vergleich zum passiven Follow-up?
 3. Ist die Zusammensetzung der Zielpopulation, insbesondere die Altersstruktur von der Art, daß die Applikation des durch Nachsorge Möglichen auch individuell durchführbar ist?
 4. Existieren Instrumente zur Messung der Effektivität?
-

Je geringer die Chance für einen erfolgreichen Eingriff in den Krankheitsprozeß oder die Relevanz des medizinisch möglichen Eingriffs einzuschätzen ist, desto geringer wird die Notwendigkeit bzw. Bereitschaft zur intensiven diagnostischen Überwachung. Maß-

geblich dafür sind die Eigenschaften der Erkrankung, z.B. Empfindlichkeit gegen Chemotherapie. Die Weitmaschigkeit einiger bestehender Nachsorgeempfehlungen wird teilweise explizit mit der ungünstigen Prognose auch früh erkannter Rezidive begründet. Aus der klinischen Erfahrung wird abgeleitet, welche Formen von Tumorprogression bei welchen Tumorarten überhaupt aussichtsreich behandelbar sind. Daß es sich überwiegend um solche Ereignisse handelt, nach denen durch intensivere apparative Nachsorge gefahndet wird, bedarf keiner Erläuterung.

In den seltensten Fällen ist jedoch a priori zu folgern, daß durch gezielte Früherkennung der Progression wirklich Vorteile im Vergleich zur exspektativen Nachsorge, d.h. erst bei Symptomen einzusetzender Behandlung resultieren. Selbst für die bereits jahrzehntelang praktizierten Krebsvorsorgeuntersuchungen ist dies nur lückenhaft belegt.

Der dritte bestimmende Faktor ist die Struktur der Zielgruppe, insbesondere die Altersstruktur. Je älter beispielsweise der Patient, desto geringer ist die Wahrscheinlichkeit zu veranschlagen, daß ein im Rahmen eines Nachsorgeprogramms erhobener positiver Befund auch tatsächlich eine therapeutische Intervention nach sich zieht. Zögert man aber die Behandlung bis zum Auftreten nicht mehr tolerabler Symptome hinaus, so war die Früherkennung durch Nachsorge letztlich ohne Wert. Auch die Verteilung prognostisch relevanter Faktoren ist zu berücksichtigen. Wenn Tumornachsorge nur für einige prognostisch günstige Stadien Erfolg verspricht, so hängt der Gesamtnutzen von der Häufigkeit dieser Stadien in der Zielpopulation ab.

All diese Bedingungen sind vor Aufnahme eines Nachsorgeprogramms nicht oder nur lückenhaft zu ermitteln. Instrumente zur Messung der Effektivität sind daher integraler Bestandteil eines Nachsorgeprogramms. Damit ist die Frage des Nutzens von Nachsorge weniger als Planungsproblem, sondern überwiegend als Bewertungsproblem charakterisiert.

2.2.2 Mit der Nachsorge verknüpfte Risiken

Dem medizinischen Nutzen, der von Nachsorgeprogrammen erwartet wird, stehen sichere und weitere mögliche Beeinträchtigungen der Teilnehmer gegenüber. Die Planung einer Nachsorgestrategie, d.h. die konkrete Ausgestaltung eines Nachsorgeprogramms mit der Festlegung von Tests und zugehörigen Frequenzen, muß zwischen dem erwarteten (erhofften) Nutzen und diesen Beeinträchtigungen abwägen.

Hierbei ist vom "Primum nil nocere" auszugehen. Mit der Definition von "nicht" beginnen jedoch bereits Probleme. Naheliegender wäre die Gleichsetzung "nicht zu schaden" = "niemandem somatische Nachteile zuzufügen". Nun ist aber bekannt, daß diagnostische Tests den betroffenen Patienten allein auf Grund ihrer Anwendung mehr oder weniger stark beeinträchtigen. Radiologische und nuklearmedizinische Untersuchungen belasten durch - wenn auch geringfügige - Strahlendosen, die Colonoskopie erfordert die Darmentleerung und verursacht damit verbundene Irritationen des Patienten, um nur einige Beispiele zu nennen. Jeder (richtig oder falsch) negative Testausgang ist so gesehen ein Schaden für den Patienten, da der negative Test keinen Anlaß gibt, anschließend zum Nutzen des Patienten zu intervenieren.

Sicher entspricht diese enge Auslegung nicht dem Sinn des Postulats. Vor allem ist zu berücksichtigen, daß die Teilnahme an der Nachsorge freiwillig ist und jeder Teilnehmer wegen der erhofften Vorteile im ungünstigen Fall bereit ist, die durch die Untersuchungen verursachten Nachteile bei negativem Testausgang, d.h. wenn sich die Durchführung der Tests ex post als unnötig erweist, in Kauf zu nehmen. Dies allein bietet jedoch keine hinreichende Basis, da Nutzen und Risiken nur unscharf prognostizierbar und dem Patienten a priori nicht bekannt sind.

Der Patient vertraut der Empfehlung seines Arztes und nimmt an, daß dieser die für ihn optimale Abwägung getroffen hat. Diese bietet den Schlüssel zur Interpretation des Postulats, nicht zu schaden. Ein Nachsorgeprogramm schadet genau dann nicht, wenn der Nutzen, den ein Patient bei Progredienz seiner Erkrankung aus der frühzeitigen Diagnosestellung zieht, die Summe der sicheren und - unter Berücksichtigung der jeweiligen Wahrscheinlichkeiten - möglichen durch die Nachsorge bedingten Nachteile aufwiegt.

In der Regel ist jedoch auch der behandelnde Arzt mit dieser Abwägung überfordert, da auch er nicht über die erforderlichen Daten verfügt. Dies führt zu der Notwendigkeit, allgemeine Nachsorgestrategien zu definieren und dem einzelnen Arzt bzw. Patienten im Rahmen von Nachsorgeprogrammen zur Verfügung zu stellen. Damit wird die kritische Abwägung von erwartetem Nutzen und möglichen Nachteilen vom behandelnden Arzt auf den planenden Experten verlagert.

Tab. 2.2: Fragen zur Einschätzung der durch ein Nachsorgeprogramm induzierten Belastungen und Risiken

1. Welche Belastungen und Risiken sind mit der Durchführung der geplanten Tests verbunden?
 2. Wie wahrscheinlich ist es, daß eine Progression diagnostiziert wird, wenn sie vorliegt (Sensitivität)?
 3. Wie wahrscheinlich ist es, daß fälschlich eine Progression vermutet wird, obwohl sie nicht vorhanden ist (1-Spezifität)?
 4. Welche Belastungen und Risiken sind mit der Abklärung falsch-positiver Befunde verbunden?
 5. Wie wahrscheinlich ist das Vorliegen des Zielereignisses (Prävalenz)?
 - 5a. Wie wahrscheinlich ist bei positivem Test das Vorliegen einer Progression (prädiktiver Wert des positiven Tests)?
 6. Sind die Wirkungen des Programms (Effizienz) meßbar?
-

Auch diesem liegen jedoch derzeit nur unvollständige Daten vor. Lückenhaft ist die Kenntnis von Sensitivität und Spezifität, den wichtigsten Indizes diagnostischer Tests, wenigstens für die geplante Anwendung der Tests im Rahmen einer ambulant durchzuführenden Tumornachsorge. Bei der Abschätzung möglicher Nachteile durch Nachsorgeuntersuchungen ist dabei vor allem der Fall zu diskutieren, daß auf Grund ungenügender Spezifität des diagnostischen Tests ein Ereignis fälschlicherweise diagnostiziert wird. In diesem Fall werden u.U. weitere Maßnahmen eingeleitet, die günstigenfalls zur Widerrufung des falsch-positiven Befundes führen. All diese Maßnahmen schlagen auf dem Konto der Nachteile oder 'Schäden' zu Buche.

Weit ungewisser noch ist jedoch in der Regel die Kenntnis des natürlichen Krankheitsverlaufs, d.h. der Häufigkeit, in der bestimmte Ereignisse zu bestimmter Zeit nach der Primärtherapie erwartet werden müssen. Man spricht von der Prävalenz. Insbesondere die wichtige Information, mit welcher Wahrscheinlichkeit das gesuchte Krankheitsereignis wirklich vorliegt, wenn ein Test angesprochen hat, der sogenannte prädiktive Wert des positiven Tests, hängt maßgeblich von der Prävalenz ab. Ohne ausreichende Daten über die Metastasierungsmuster und damit die Prävalenz von Ereignissen im zeitlichen Verlauf besteht für die Planung von Nachsorgestrategien das Risiko, an der Realität vorbeizulaufen.

Bezüglich der Sensitivität und Spezifität diagnostischer Tests im Rahmen von Nachsorgeprogrammen sowie der Prävalenz von Zielereignissen ergeben sich besondere Anforderungen und spezifische Probleme, die im 3. Abschnitt behandelt werden. Anhand von Modellrechnungen wird im 6. Abschnitt zusätzlich aufgezeigt, wie vorhandene Daten im Hinblick auf die Nachsorgeplanung nutzbar gemacht werden können.

2.2.3 Probleme der Finanzierbarkeit

Tumorzentren stehen bei der niedergelassenen Ärzteschaft gelegentlich im Verdacht, Patienten systematisch in eigenen Ambulanzen zu betreuen und damit Funktionen zu übernehmen, die der vereinbarten Arbeitsteilung zwischen Kliniken und Kassenärzten widersprechen. Dieser Vorwurf ist in der Regel nicht gerechtfertigt. Er ist aber als Beleg dafür zu werten, daß Dienstleistungen, die im Rahmen der Tumornachsorge erbracht werden, für die niedergelassenen Ärzte einen ökonomisch bedeutsamen Faktor darstellen können. Die Systematisierung der Nachsorge durch Nachsorgeprogramme wird ihre ökonomische Bedeutung weiter erhöhen.

Zur oberflächlichen Orientierung über den ungefähren Rahmen des ökonomischen Aufwands bei maximalem Versorgungsangebot im Rahmen der Tumornachsorge kann die Summe von 395 Mio. DM pro Jahr, die derzeit für die Krebsvorsorge aufgebracht wird (Zit. nach (32)), als Vergleich herangezogen werden. Der potentielle Teilnehmerkreis für Nachsorgeprogramme ist zwar mit ca. 1% der Bevölkerung niedrig. Dieser Wert ergibt sich aus einer Inzidenz von ca. 400

pro 100.000 und einer 5-Jahres-Überlebensrate von etwa 40%. Andererseits jedoch sind für einige publizierte Nachsorgestrategien Kosten von 300 DM und mehr pro Untersuchung nicht zu hoch gegriffen, wobei diese mehrfach pro Jahr anfallen können. Zusätzlich liegt die Wahrscheinlichkeit für befundabhängige weitere Diagnostikmaßnahmen in der Nachsorge höher als bei Krebsvorsorgeuntersuchungen. Damit könnten die Kosten pro Einzelfall und Jahr in der Nachsorge die Größenordnung von 1.000 DM erreichen. Unter der Annahme, daß etwa 60-70% der Tumorpatienten das Angebot der Nachsorgeprogramme wahrnehmen würden, wären ca. 360.000 Patienten zu versorgen. Aus dem Produkt dieser Fallzahl mit dem Aufwand pro Fall und Jahr ergäbe sich ein der Krebsvorsorge vergleichbarer Gesamtaufwand.

Bei einer allgemeinen Einführung von engmaschigen und apparativ aufwendigen Tumornachsorgeprogrammen würden somit erhebliche Kosten entstehen. Die Rate behandlungsbedürftiger, nicht tumorbedingter Nebenbefunde, die den Patienten ohne Tumornachsorge nicht zum Arzt geführt hätten, ist hierbei noch zu den direkten Aufwendungen hinzuzurechnen. Bei der Krebsvorsorge wird diese als hoch eingeschätzt (1). Es besteht kein Grund zu der Annahme, daß bei den im Durchschnitt älteren Patienten in der Tumornachsorge weniger Nebenbefunde anfallen könnten. Nicht zuletzt auch wegen dieser volkswirtschaftlichen Konsequenzen ist bei der Planung von Nachsorgeprogrammen Sorgfalt gefordert.

2.3 Probleme der Evaluation von Nachsorgeprogrammen

Die Trennung von Planung und Bewertung eines Nachsorgeprogramms ist bis zu einem gewissen Grad arbiträr, da Planung, Evaluation, Modifikation gerade bei anfänglicher Unsicherheit verzahnt und aufeinander abgestimmt werden müssen, wenn Nachsorgeprogramme auch aus methodischen und organisatorisch-technischen Gründen nach ihrer Einführung über eine gewisse Zeit nicht verändert werden sollen und können. Teile dieses iterativen Prozesses von Planung, Evaluation und Modifikation könnten evtl. mit Hilfe von Simulationstechniken antizipiert werden. Dies würde zu noch engerer Verzahnung der Teilaspekte Planung und Bewertung führen. Die Kernfrage, die im vorangegangenen Abschnitt zurückgestellt wurde, bleibt jedoch als klares Bewertungsproblem bestehen: Profitiert der Patient durch Nachsorge?

2.3.1 Bewertungskriterien für Nachsorgeprogramme

Nach der Definition in Abschnitt 2.1.2 umfaßt ein Nachsorgeprogramm medizinische, organisatorische, ökonomische und datentechnische Komponenten. Aus jedem dieser Bereiche können Fragen an Nachsorgeprogramme formuliert werden. Grundlegend sind die Fragen nach Wirksamkeit und Wirkungen, also

welcher Nutzen resultiert aus der Teilnahme an Nachsorgeprogrammen?

mit welchen Risiken ist die Teilnahme verbunden?

Diese Fragen sind allen anderen Fragen übergeordnet. Wenn der Nutzen eines Nachsorgeprogramms nicht belegt werden kann, verlieren alle anderen Fragen ihren Sinn. Der Begriff 'Nutzen' resultiert hierbei aus einer Abwägung von Chancen im Fall einer Progression, 'unnötiger' Beeinträchtigung im Fall des negativen Testausgangs und Belastungen durch die Abklärung falsch-positiver Befunde. Es ist zu berücksichtigen, daß damit noch nichts über potentielle Indikatoren gesagt ist, an Hand derer diese Abwägung im Einzelfall getroffen werden kann.

Wird die Frage nach dem Nutzen - auf welcher Grundlage auch immer - bejaht, so stellt sich die nächste Frage

ist durch die Organisation gewährleistet, daß ein hoher Anteil der Zielgruppe einen hohen Prozentsatz des erreichbaren Nutzens zieht?

Auch diese bereits in der Formulierung komplexe Frage fächert sich in zahlreiche heterogene Einzelaspekte auf, z.B. die Teilnahmequote für verschiedene Schichten der Zielgruppe.

Bereits mehrfach wurde zum Ausdruck gebracht, daß der von Nachsorgeprogrammen ausgehende Nutzen aus der Abwägung von Chancen und Risiken resultiert, wobei im Einzelfall im vorhinein nicht bekannt ist, welcher der möglichen Fälle gegeben ist. M.a.W. handelt es sich vorwiegend um einen Populationsnutzen, der auf Grund unterschiedlicher Teilnahmequoten innerhalb verschiedener Untergruppen unterschiedliche Werte annehmen kann. Aus der Analyse verschiedener Screeningprogramme ist bekannt, daß Gruppen mit hohem Risiko das jeweilige Angebot seltener wahrnehmen (1, 5, 10, 58).

Weitere Aspekte zur Bewertung der Organisation eines Nachsorgeprogramms sind die Kommunikation der Ärzte und Institutionen über einzelne Patienten, Ausstattung der beteiligten Institutionen und Qualifikation der durchführenden Ärzte.

Ein anderes Licht auf Nachsorgeprogramme wirft die folgende Frage:

rechtfertigt der Nutzen des Nachsorgeprogramms die entstandenen Kosten?

Wirtschaftliche Anstrengungen und menschliches Leben bzw. menschliche Gesundheit sind inkommensurable Größen. Damit ist diese Frage eine Wertfrage, die durch die Aufnahme eines Nachsorgeprogramms bereits teilweise beantwortet ist. Zur Bewertung könnte stattdessen die Frage gestellt werden:

haben sich die tatsächlich anfallenden Kosten im projektierten Rahmen gehalten?

Welche Probleme jedoch allein die Schätzung der tatsächlichen Kosten aufwerfen kann, wird für das Beispiel der Krebsfrüherkennung exemplarisch aus (69) deutlich.

Die letzte Frage schließlich liegt auf einer anderen Ebene:

werden die durch das Nachsorgeprogramm verursachten Effekte hinreichend transparent, so daß eine Beantwortung der vorhergehenden Fragen möglich wird?

Eine verneinende Antwort auf diese Frage impliziert die Unmöglichkeit der Bewertung des gesamten Nachsorgeprogramms. Zurückbliebe dann das Unbehagen der Beteiligten, ein Instrument zu verwalten, dessen wirklicher Nutzen fragwürdig ist. In dieser Situation befinden sich einige der Krebsfrüherkennungsmaßnahmen in der Bundesrepublik. Beispielsweise für das Screening der Prostata ist zu beobachten, daß kaum noch vorbehaltlos zur Teilnahme aufgefordert wird. Sinkende Inanspruchnahme, so wurde im Geleitwort zu der 1983 durchgeführten Studie des BMFT über Krebsvorsorge (zit. nach (32)) argumentiert, könnte die Frage der Weiterführung des derzeitigen Programms bald beantworten. Auch an anderen Stellen ist festzustellen, daß die Diskussion um Vorsorgeuntersuchungen nicht immer emotionslos geführt wird. So wird in (14) das Infragestellen solcher Programme als Modeerscheinung apostrophiert, in (16) fallen entschiedene Äußerungen über den Stellenwert der Krebsvorsorge, um nur Beispiele zu nennen. Im Bereich der Tumor-

nachsorge müssen alle Anstrengungen unternommen werden, daß sich Vergleichbares nicht wiederholt und die Diskussion sich auf eine transparente Faktenbasis stützen kann.

2.3.2 Studienansätze

Jede Analysetechnik zur Evaluation von Nachsorgemaßnahmen erfordert Daten. Bei der Planung von Studien entscheidet sich, ob letztlich die zur Analyse erforderlichen Daten in der erforderlichen Qualität verfügbar sein werden. Was bei der Studienplanung und folglich der Gewinnung von Daten versäumt wurde, läßt sich später auch durch differenzierte Auswertungstechniken nur begrenzt kompensieren. Andererseits muß davon ausgegangen werden, daß selbst bei optimaler Planung eines Nachsorgeprogramms nicht alle Daten verfügbar werden, die zur Evaluation erforderlich wären. Die folgende kurze Überlegung soll dies verdeutlichen.

Die HIP-Studie ist eine der wenigen und zugleich die bekannteste und vielleicht auch am besten geplante randomisierte Studie über den Wert von Früherkennungsmaßnahmen. Studienziel war die Beantwortung der Frage, ob periodisches Screening mit Mammographie und klinischer Untersuchung der Brust die Mortalität an Brustkrebs senken kann. Hierzu wurden aus den Reihen der Mitglieder des HIP (Health Insurance Plan) zwei Zufallsstichproben von je 31.000 Frauen zwischen 40 und 64 Jahren gezogen. Den Mitgliedern der Studiengruppe wurden jährliche Screeninguntersuchungen angeboten, während die Mitglieder der Kontrollgruppe ihre Gewohnheiten bezüglich Inanspruchnahme medizinischer Leistungen beibehielten (57). Die Studie erbrachte eine signifikante Reduktion der Mortalität am Mamma-Ca in der Screening-Gruppe.

Wegen der unterschiedlichen Kosten und Risiken der simultan benutzten Tests wurde mehrfach versucht, den relativen Beitrag von Mammographie bzw. Tastbefund zum positiven Gesamtergebnis zu ermitteln (z.B. (74)). Diesen Ansätzen wurde jedoch entgegengehalten, daß die Extrapolation des Gesamtergebnisses auf einzelne Untergruppen, z.B. auf die Untergruppe der durch Tastbefund diagnostizierten Fälle, nur unter Annahmen möglich sei, die zwar plausibel, nicht aber zwingend sind. Streng genommen beruht die retrospektive Stratifikation der entdeckten Fälle nach Tastbefund

oder Mammographie auf einer Variablen, deren Wert sich erst nach der Randomisierung ergibt bzw. die nicht für alle Patienten in der Studie bestimmt wird.

Solche Kriterien, die auch als 'Pseudovariablen' bezeichnet werden (48), sind auf Grund ihrer Genese von limitiertem Wert in der statistischen Analyse. In der HIP-Studie wurde daher durch die Randomisation vorwiegend die Gesamtaussage unterstützt, betreffend die Anwendung des definierten Programms auf eine Zielpopulation der entsprechenden Zusammensetzung. Schon die wichtige Interpretation von Ergebnissen in Untergruppen ist nicht mehr durch die Randomisation gestützt.

Diese Überlegung stellt nicht den Wert randomisierter Studien in Frage. Randomisierte Studien sind der einzige Weg, der die Prüfung von Hypothesen mit definierten Irrtumswahrscheinlichkeiten gestattet, wenn das Instrumentarium korrekt angewandt wurde. Gleichzeitig sind sie auch der aufwendigste Weg, sich mit einer Hypothese auseinanderzusetzen. Vor der Durchführung randomisierter Studien ist deshalb u.a. zu prüfen, ob Aufwand und Ergebnis in ausgewogenem Verhältnis stehen.

In der Tumornachsorge liegen multiple, konkurrierende Zielereignisse vor, wesentlich mehr als ein oder zwei Tests werden simultan, verzahnt oder sequentiell eingesetzt. Die vorrangige Frage bezieht sich nicht auf Durchführung eines bestimmten Nachsorgeprogramms vs. Unterlassung von Tumornachsorge schlechthin, sondern auf differenzierte Abstufungen und damit Optimierung der Strategie. Wann soll welches Organ durch welche Tests kontrolliert werden? Selbst ohne Berücksichtigung von Untergruppen in der Zielpopulation hat man es mit einer Vielzahl simultan auftretender Einzelfragen zu tun, denen man nicht gleichzeitig durch randomisierte Designs gerecht werden könnte. Durchführbar wären Ansätze, in denen ein Faktor kontrolliert verändert wird, während andere Faktoren konstant gehalten werden, oder Studien, in denen zwei unterschiedliche Strategien verglichen werden.

Verzichtet man auf Randomisation, so verliert der Versuch, zwischen Nachsorgemaßnahmen und beobachteten Unterschieden im Zielkriterium einen Kausalzusammenhang herzustellen, an Aussagekraft.

Die Evaluation ist dann auf andere, weniger aussagekräftige Verfahren angewiesen. Um hierbei möglichst differenzierte Techniken anwenden zu können, muß u.a. die Erfassung möglichst vieler prognostisch relevanter Faktoren angestrebt werden.

Die Festlegung der zur Bewertung von Nachsorgeprogrammen 'erforderlichen' Daten stellt sich in diesem Fall größtenteils als Optimierungsproblem. Es geht darum, realistische Konzepte zu entwickeln, die mit vertretbarem Aufwand praktiziert werden können und durch die die anstehenden Fragen auf einem akzeptablen Wahrscheinlichkeits- bzw. Plausibilitätsniveau beantwortbar werden. Methodischer Rigorismus könnte dabei ebenso verfehlt sein wie die Vernachlässigung methodischer Anforderungen im Hinblick auf die spätere Bewertung der Maßnahmen. 'Methodik' beginnt und endet nicht bei der Randomisation. Im 4. Abschnitt werden Wege zwischen Extrempositionen aufgezeigt.

2.4 Analogien zwischen Screening und Tumornachsorge

Auf Screeningprogramme und die daraus gewonnenen Erfahrungen wurde bereits mehrfach Bezug genommen. Dies unterstreicht, daß Analogien zwischen Screening, speziell der Krebsvorsorge, und der Nachsorge an Tumorkranken bestehen.

Sowohl Screening als auch Nachsorge richten sich an einen asymptomatischen Personenkreis, beide streben die frühzeitige Entdeckung eines als progredient angesehenen Krankheitsherd an.

Beim Screening ist das Auftreten des Zielereignisses selten und daher als unerwartet einzustufen. Demgegenüber sind positive Befunde in der Nachsorge zwar auf Grund der bekannten Dynamik der Krebskrankheit durchaus erwartet. Dennoch arbeitet auch die diagnostische Nachsorge in niedrigen Prävalenzbereichen, da hier die Untersuchungen engmaschiger durchgeführt werden als beim Screening. In beiden Fällen konkurrieren Frühzeitigkeit der Entdeckung und Anreicherung der Prävalenz. Je frühzeitiger ein Ereignis erfaßt werden soll, desto engmaschiger sind die entsprechenden Kontrollen durchzuführen. Daraus folgt ein Rückgang der pro Untersuchung sich ergebenden Prävalenz.

Sensitivität, Spezifität und Inzidenz des Zielereignisses sind daher sowohl für Screening als auch in der Tumornachsorge planungsrelevant. Dennoch kann nicht a priori davon ausgegangen werden, daß aus gleichen Zahlenverhältnissen, etwa gleichem prädiktiven Wert des positiven Tests, für Screening und Nachsorge gleiche Schlüsse zu ziehen sind.

Analogien ergeben sich auch im Hinblick auf die jeweils geeigneten methodischen Ansätze zur Evaluation des medizinischen Nutzens. Die HIP-Studie als Beispiel einer großen randomisierten Studie zur Bewertung eines Krebsfrüherkennungsprogramms wurde bereits genannt. Daneben liegen Ergebnisse aus zahlreichen Fall-Kontroll-Studien vor, in denen das aus der Teilnahme vs. Nichtteilnahme an der Krebsfrüherkennung resultierende relative Risiko geschätzt wird. Diese und weitere Studien gestatten es, die Methodendiskussion bezüglich der Evaluation des medizinischen Nutzens von Screeningmaßnahmen an Hand konkreter Beispiele und empirischer Daten zu führen. U.a. wird es möglich, die mit unterschiedlichen methodischen Ansätzen gewonnenen Ergebnisse zu vergleichen und auf Konsistenz zu überprüfen, z.B. (1, 33).

Zum Stellenwert von Tumornachsorgeprogrammen wurden vergleichbare Studien bisher nicht durchgeführt. Daher ist in der Methodendiskussion über die Evaluation von Nachsorgeprogrammen der Rückgriff auf die im Screening gewonnenen Erfahrungen besonders vorteilhaft. Ob und in welchem Umfang die im Screening gefundenen methodischen Probleme und Lösungen auf die Bewertung von Nachsorgeprogrammen übertragen werden können, kann jedoch nur theoretisch, d.h. losgelöst von empirischen Daten, diskutiert werden.

Analogien ergeben sich schließlich auch bezüglich der sozialmedizinischen und ökonomischen Bewertung von Screening und Nachsorge. Diese verläßt zwar den Rahmen der statistisch epidemiologischen Methodik, sollte aber von methodisch einwandfreien Ergebnissen bestimmt sein.

Von der WHO wurde im Jahre 1968 ein 10-Punkte-Katalog zur Beurteilung von Screeningprojekten und -programmen aufgestellt (71), der die genannten Bereiche Planung und medizinische Bewertung sowie sozialmedizinische bzw. ökonomische Begründung abdeckt.

Diese Zusammenstellung wird bis heute weitgehend unverändert als Grundlage für Planung und Bewertung von Screeningprogrammen akzeptiert. Bisher wurde nicht untersucht, ob und inwieweit die WHO-Kriterien auf die Situation in der Tumornachsorge übertragbar sind.

2.5 Derzeitige Situation der Tumornachsorge

2.5.1 Tumornachsorge an Kliniken und Tumorzentren

Seit 1977 wurden in der Bundesrepublik etwa 30 Tumorzentren gegründet. In der Regel wird die Leitfunktion eines Tumorzentrums für die umliegende Region u.a. dadurch zum Ausdruck gebracht, daß es Nachsorgestrategien anbietet, die jeweils als medizinisch notwendig angesehen werden. Nicht alle derartigen Schemata und Empfehlungen sind über die Literatur zugänglich, die Publikation erfolgt vielfach im Selbstverlag, wie z.B. auch die Manuale des Tumorzentrums München. Speziell zum Mammakarzinom sollen in den letzten Jahren in der Bundesrepublik über 40 verschiedene Nachsorgeempfehlungen erarbeitet worden sein. Aus diesen Aktivitäten wird erkennbar, daß hier einem akuten, rasch wachsenden Bedarf entsprochen wird, ein existierendes Handlungsbedürfnis verlangt nach Entscheidungshilfen.

Diese Vielfalt von Empfehlungen deutet aber auch darauf hin, daß die verfügbaren Daten und Informationen zur Planung und Begründung optimaler Nachsorgestrategien offenbar nicht ausreichen. Je schmaler die allgemein zugängliche Datenbasis, desto stärker fließt die individuelle Erfahrung in die Definition von Strategien ein. Diese Erfahrung hängt von Zufälligkeiten ab, besonders bei selteneren Ereignissen, und ist kein zuverlässiger Spiegel der Wirklichkeit. Dies erklärt teilweise die beobachtbare Vielfalt der Empfehlungen.

Das Angebot von telefonischen Tumorkonsilien (z.B. im Tumorzentrum München) für umliegende Krankenhäuser und Kassenärzte formalisiert in gewisser Hinsicht auch den Austausch von individuellen Vorgehensweisen und Erfahrungen. Zusätzlich bietet es den ange-

sprochenen Ärzten die Möglichkeit, Verunsicherung abzubauen, die beim Studium widersprüchlicher Nachsorgeempfehlungen verschiedener kompetenter Stellen unvermeidlich sein dürfte.

Die von Tumorzentren erarbeiteten Strategien zur Nachsorge klinisch tumorfreier Patienten haben den Charakter von Empfehlungen. Den Urhebern dieser Empfehlungen wird hierbei nicht unbedingt die spätere kritische Reflexion ermöglicht oder auch abverlangt. Auch die Frage, in welchem Prozentsatz diesen Empfehlungen im jeweiligen Einzugsgebiet nachgekommen wird, ist noch nicht untersucht worden, da eine systematische Erfassung relevanter Daten zur Bewertung der erarbeiteten Strategien z.Z. noch nicht erfolgt.

2.5.2 Die Nachsorgeprogramme der Kassenärztlichen Vereinigung Bayerns

In Bayern wurden für die Tumoren der Cervix, des Endometriums und der Mamma im Jahre 1979 von der Kassenärztlichen Vereinigung erstmals in der Bundesrepublik Nachsorgeprogramme eingeführt (35). Leistungen, die im Rahmen dieser Programme erbracht werden, können vom Kassenarzt gesondert abgerechnet werden. An diese Abrechnung ist die Verpflichtung zur Dokumentation der erhobenen Befunde gekoppelt.

Längsschnittanalysen dieser Daten sind bedingt möglich, das Record-Linkage erfolgt über die Nummer eines tumorspezifischen Nachsorgekalenders. Die Identifikation der Trägerin anhand der Kalendernummer kann nur bei der Institution erfolgen, die der Patientin den Kalender ausgehändigt hat. Insofern verbindet sich die Notwendigkeit der Zuordenbarkeit von Folgebefunden mit der Erfordernissen der anonymisierten Datenspeicherung.

An dieser sog. 'programmierten Nachsorge' der Kassenärztlichen Vereinigung Bayerns wurde verschiedentlich auch Kritik geübt. Bisher liegt noch keine Auswertung vor, die Ansätze zur Bewertung der Programme bietet.

2.5.3 Tumornachsorge im Einzugsbereich des Tumorzentrums München

Im Tumorzentrum München wurden für die wichtigsten Tumorarten Projektgruppen gegründet, zu deren Aufgaben die Erarbeitung von Empfehlungen zur Diagnostik, Therapie und Nachsorge des betreffenden Tumors zählt. Zum Teil existieren Empfehlungen des Tumorzentrums München und der Kassenärztlichen Vereinigung Bayern nebeneinander.

Seit Anfang 1982 wird im Tumorzentrum München ein allgemeiner Nachsorge-Terminkalender ausgegeben (26, S. 26 ff). Für die Planung und Bewertung von Nachsorgestrategien stehen damit verbundene organisatorische Vorstellungen im Vordergrund (25). Wie die tumorspezifischen Nachsorgekalender der Kassenärztlichen Vereinigung Bayerns trägt auch der Nachsorge-Terminkalender des Tumorzentrums München eine eindeutige Nummer. Wie bei den Kalendern der Kassenärztlichen Vereinigung wird die Zuordnung von Identität des Trägers und Nummer des Kalenders vom Aussteller des Kalenders festgehalten. Für eine Zuordnung von Informationen zum Krankheitsverlauf würde es genügen, wenn jeder Kassenarzt die Kalendernummer in Verbindung mit den erhobenen Befunden mitteilt. So könnte sich ein Informationsverbund ergeben, der dem von der Kassenärztlichen Vereinigung praktizierten Vorgehen nahekam. Zusätzlich bestünde die Chance, durch die Koppelung mit einem Tumorregister zur Information aus der Nachsorge auch verlaufsbestimmende charakteristische Daten über Primärbefund und -therapie verfügbar zu haben.

2.5.4 Das Tumorregister München

Seit 1978 baut das Tumorzentrum München eine klinikübergreifende Verlaufsdokumentation, das Tumorregister München, auf. Basisdaten über Primärbefund, Behandlung und, soweit verfügbar, Informationen zum Krankheitsverlauf werden im Register erfaßt. Ein funktionierendes Tumorregister ist das Fundament der kontinuierlichen Langzeitbeobachtung von Tumorpatienten. Aber auch für die spätere Evaluation der Nachsorge für die Patienten des Tumorzentrums München kann es als Basis dienen.

Am 31.12.83 waren im Tumorregister München ca. 23.000 Patienten erfaßt. Tab. 2.3 zeigt auszugsweise die Verteilung der dokumentierten Diagnosen. Diese hängt von der Auswahl der kooperierenden Kliniken ab und spiegelt nicht die Inzidenzverhältnisse in Südbayern wider. Die organisatorische und technische Leistungsfähigkeit des Registers wird jedoch durch diese Fallzahl belegt, auch unabhängig von Repräsentativität.

Tab. 2.3: Fallzahlen der häufigsten der im Tumorregister München erfaßten Diagnosen, Stand 31.12.1983

	n	%
Larynx	683	3.06
Magen	819	3.67
Dickdarm	1614	7.24
Lunge/Bronchien	1361	6.10
Malignes Melanom	2160	9.68
Mamma	4391	19.68
Cervix	2325	10.42
Endometrium	1932	8.66
Ovarien	1013	4.54
Malignes Lymphom	880	3.94
Gesamtfallzahl		
(alle Diagnosen)	22317	100.00

Auf Schwierigkeiten stößt die Vervollständigung des Follow-up. Über Patienten, die nicht von einer mit dem Tumorregister kooperierenden Klinik nachgesorgt werden, erhält das Register in der Regel keine Verlaufsinformationen. Nicht nur die Verteilungen der Anzahl von Verlaufsinformationen bieten daher ein heterogenes Bild. Verlaufsauswertungen sind derzeit für die meisten Tumordiagnosen nur für einen Ausschnitt der erfaßten Fälle möglich.

3. Methodische Anforderungen bei der Planung von Nachsorgeprogrammen

3.1 Präzision diagnostischer Tests

3.1.1 Ziele und Anwendungen diagnostischer Tests

3.1.1.1 Diagnostik zur Entscheidungsfindung

Häufigstes Ziel der Diagnostik ist die Diagnosestellung, das ist "der Prozeß, der zur Erkennung einer Klasse oder Gruppe führt, der eine Krankheit angehört, so daß die nachfolgenden Behandlungsmaßnahmen den Nutzen des Patienten, ausgehend vom verfügbaren Wissen, maximieren" (51).

Hierbei ist der Begriff 'Diagnose' offengelassen. Sowohl 'Intrathorakale Geschwulst' als auch 'technisch operables, jedoch funktionell inoperables peripheres Karzinom des rechten Oberlappens von großzelligem Typ' sind, je nach Zweckbestimmung, sinnvolle Diagnosen. Unter 'Diagnosestellung' kann daher auch die Präzisierung einer Diagnose verstanden werden, z.B. die Abklärung der Operabilität eines Karzinoms.

Tab. 3.1: Grundmuster des Vorgehens bei der Anwendung diagnostischer Tests zur Diagnosestellung (nach 51)

Strategie	Bedeutung
Pattern recognition	das (i.a. unmittelbare) Erkennen, daß die Beschwerden des Patienten einem bekannten Krankheitsbild entsprechen
Entscheidungsbaum	Schrittweises diagnostisches Vorgehen, in dem jeweils der Ausgang einer Untersuchung den nächsten Schritt bestimmt
Ausschöpfung	Parallele Beobachtung zahlreicher Fakten und Befunde, anschließende Sichtung der Ergebnisse
hypothesengesteuert - deduktiv	Formulierung einer 'kurzen Liste' von möglichen Diagnosen auf Grund des ersten Eindrucks über den Patienten, anschließend Ausschöpfung derjenigen diagnostischen Tests, die am besten zur Reduktion dieser Liste geeignet sind (Differentialdiagnostik)

Hilfsmittel zur Diagnosefindung sind geeignete Untersuchungsverfahren, die Aufschluß über den Zustand des Patienten geben können, sog. diagnostische Tests. Die Schwierigkeit der Einengung dieses Begriffs wird gerade am Beispiel 'pattern recognition' ersichtlich. Im Begriffsumfang sicher enthalten sind Laboruntersuchungen und Verfahren der apparativen Diagnostik, wie z.B. Röntgen oder/und Ultraschall. Daneben kann es sinnvoll sein, auch Befunde, Symptome oder sogar anamnestische Angaben hinzuzurechnen, vor allen Dingen, wenn sie als 'pattern' in Erscheinung treten (23). Entscheidend ist, ob unter der jeweiligen Fragestellung aus der Auswertung der entsprechenden Informationen Rückschlüsse auf die Diagnose ermöglicht bzw. erleichtert werden. Für die formale Bewertung der Richtigkeit diagnostischer Schlußfolgerungen, der Folgen von Fehleinschätzungen oder des Stellenwerts bestimmter Untersuchungen in einem diagnostischen Gesamtkonzept spielt es im Prinzip keine Rolle, auf welche Weise diagnostische Information gewonnen wurde.

Die Grundmuster der Tabelle 3.1 können zu unterschiedlichen Diagnostikstrategien zusammengesetzt werden. Tests können parallel (im Sinne der 'Ausschöpfung') oder sequentiell (im Rahmen eines 'Entscheidungsbaums') benutzt werden, sie können dem Ausschluß von Möglichkeiten (Verkürzung der 'kurzen Liste') oder der Bestätigung von Vermutungen ('Pattern recognition') dienen. Eine Diagnose kann, ähnlich wie ein mathematischer Beweis, auf unterschiedlichem Weg zustandekommen. Sie kann entweder auf direktem Weg oder durch Ausschluß anderer Möglichkeiten gewonnen werden oder sich erst durch eine Kombination beider Wege ergeben. Unterschiedliche Fragestellungen erfordern unterschiedliche diagnostische Strategien. Die Bedeutung eines bestimmten diagnostischen Tests kann in verschiedenen Umgebungen unterschiedlich sein.

Nicht jeder Test ist in gleicher Weise zur Bestätigung und zum Ausschluß von Vermutungen geeignet. Zum Ausschluß einer Diagnose wird ein Test benötigt, der auf die betreffende Krankheit mit hoher Wahrscheinlichkeit anspricht, wenn sie vorliegt, d.h. dessen Sensivität hoch ist. Aus dem negativen Ergebnis eines hochsensitiven Tests kann im Prinzip auf die Abwesenheit der in Frage stehenden Krankheit geschlossen werden.

Auf das Vorliegen einer Krankheit kann im Prinzip geschlossen werden, wenn ein Test anspricht, der auf andere in Frage kommende Krankheiten nur selten ein positives Ergebnis zeigt, d.h. dessen Spezifität hoch ist.

Die Sensitivität beschreibt das Verhalten des Tests in der Gruppe der Erkrankten und wird durch die Interaktion von Test und Krankheit bestimmt, während die Spezifität als Aussage über die nicht Erkrankten vom jeweiligen Angebot an Krankheitsalternativen abhängt.

3.1.1.2 Diagnostik im Screening

Diagnostikverfahren werden nicht nur zur Diagnosefindung am Kranken eingesetzt. Auch 'Gesunde' unterziehen sich in unterschiedlichen Situationen diagnostischen Tests, z.B. bei Einstellungsuntersuchungen, bei Überwachung und Kontrolle von Personen nach dem Auftreten einer infektiösen Krankheit sowie bei Reihenuntersuchungen zur Früherkennung von Krankheiten, sogenannten Screeninguntersuchungen. Krebsvorsorge, Röntgenreihenuntersuchungen und die Vorsorgeuntersuchungen an Kleinkindern sind die bekanntesten Screeningprogramme in der Bundesrepublik.

Untersuchungen an Gesunden lassen sich als 'hypothesengesteuert - deduktives' Vorgehen theoretisch in das Schema der Tabelle 3.1 einordnen. Die 'kurze Liste' von Diagnosen besteht jeweils aus zwei Elementen. Die erste Diagnose lautet 'gesund', die zweite Diagnose beispielsweise 'Mammakarzinom'. Die Screeninguntersuchung mit einem oder wenigen Tests dient dem Ausschluß der zweiten Möglichkeit, die erwartete Diagnose 'gesund' wird in der Regel bestätigt. Wenn nicht, folgen weitere Aktionen, z.B. die Abklärung der Verdachtsdiagnose über einen Entscheidungsbaum. Bei der Untersuchung von Kleinkindern im Rahmen des Früherkennungsprogramms werden simultan mehrere 'kurze Listen' abgearbeitet, z.B. 'gesund/Hüftgelenksluxation', 'gesund/Sprachstörung' usw. Solches Screening, das parallel mehrere, unabhängige Zielereignisse berücksichtigt, wird als multiphasisches Screening bezeichnet.

3.1.1.3 Diagnostik im Rahmen von Verlaufsbeobachtung

Die Abklärung eines positiven Testergebnisses läßt erkennen, ob das Testergebnis valide, der Befund berechtigt war. Nach der diagnostischen Abklärung steht die Auswahl einer geeigneten Therapie an. Zur Indikationsstellung und Steuerung der Therapie sind auch nach Diagnosestellung weitere Untersuchungen erforderlich. Hierzu werden vorwiegend Tests herangezogen, deren Validität - bezogen auf den Einzelfall - im Rahmen der Diagnosefindung sichergestellt wurde. Exemplarische Anwendungen zeigt die Tab. 3.2.

Tab. 3.2: Typische Anwendungen diagnostischer Tests zur Verlaufsbeobachtung bei bekannter Diagnose

Zeitpunkt	Ziel	Beispiel
Vor einer Intervention	Beobachtung einer Entwicklung	Erfassung des Zeitpunktes der beginnenden Behandlungsbedürftigkeit einer Hypergammaglobulinämie
	Ausschluß von Kontraindikationen	Lungenfunktionsprüfung vor Pneumonektomie
Während einer Intervention	Steuerung	Beobachtung des Blutbildes unter Chemotherapie
Nach einer Intervention	Beobachtung des Genesungsprozesses	Röntgenkontrollen nach traumatisch bedingter Fraktur

Unter bestimmten Voraussetzungen, z.B. im Rahmen einer internistischen Langzeittherapie eines metastasierenden Mammakarzinoms, sind die Stadien dieser Tabelle zyklisch bzw. wiederholt zu durchlaufen.

Nicht jede Wiederholung diagnostischer Tests im zeitlichen Verlauf ist als Verlaufsbeobachtung einzustufen. Voraussetzung der Verlaufsbeobachtung ist die Gewißheit über die Validität des Tests, bezogen auf den betreffenden Patienten. Verlaufsbeobachtung bedeutet, aus der Beobachtung von Testergebnissen im zeitlichen Verlauf auf die Veränderung der Krankheit zu schließen.

Häufig tritt dagegen die Situation auf, daß ein Test im zeitlichen Abstand gerade dazu wiederholt wird, um die Validität der

Befunde zu überprüfen. In diesem Fall wird also aus dem Vergleich von Testergebnissen auf das Vorhandensein der Krankheit geschlossen, so daß die Testwiederholung als Teil der Diagnosestellung anzusehen ist.

3.1.1.4 Diagnostik in der Tumornachsorge

Ziel von Diagnostikmaßnahmen in der Tumornachsorge können sowohl die Entscheidungsfindung als auch Screening oder Verlaufsbeobachtung sein: teils treffen diese Anforderungen beim gleichen Patienten zusammen. Planbar im Rahmen von Nachsorgeprogrammen sind nur die letztgenannten Anwendungen, also Screening und Verlaufsbeobachtung.

Zwischen Screening und Verlaufsbeobachtung sind die Übergänge in der Nachsorge fließend. Eindeutig ist die Zuordnung, wenn eine Untersuchung entweder ausschließlich der Früherkennung einer Tumorprogression oder ausschließlich der Beobachtung des Therapieergebnisses bzw. der Steuerung einer Therapie dient.

Vielfach kreuzen sich diese Anforderungen jedoch, und aus der Art einer Untersuchung (z.B. Röntgen der Lunge) allein ist die mit der Durchführung verknüpfte Absicht nicht erkennbar. Dies ist bei der Bewertung von Nachsorgeprogrammen zu berücksichtigen.

Der potentiell systemische Charakter der Tumorkrankheit stellt ein weiteres Problem. Tumorprogression kann sich an zahlreichen Orten des Körpers manifestieren, so daß sich das Zielkriterium der prophylaktischen Nachsorge in zahlreiche, voneinander abhängige Einzelkriterien auflöst. Weder die Klassifikation als monophasisches, noch die Einordnung als multiphasisches Screening würde allen Phasen der Tumornachsorge gerecht.

3.1.2 Güte-Indizes für diagnostische Tests

3.1.2.1 Prävalenzunabhängige Indizes

Viele diagnostische Verfahren liefern in der Praxis komplexe, vielschichtige Informationen, die oftmals sogar in sich heterogen oder widersprüchlich sind, wie z.B. zahlreiche bildgebende Ver-

fahren. Die Befundung solcher komplexer Testergebnisse kann beträchtliche Schwierigkeiten bereiten. Einfacher zu beurteilen sind Tests, deren Ergebnis in irgendeiner Form quantitativ ist, z.B. Blutdruck oder die meisten Labortests. Die wenigsten Tests liefern ein für Entscheidungsbäume notwendiges zweiwertiges Ergebnis 'positiv' oder 'negativ', wenn nicht weitere Verarbeitungsschritte folgen. Für die Mehrzahl der diagnostischen Tests beruht ein letztendlicher ja/nein Befund auf Datenreduktion, -interpretation und -klassifikation durch den untersuchenden Arzt.

Betrachtet man Test und zugehörige Interpretation als Ganzes, so ergeben sich die bekannten Möglichkeiten der richtigen und falschen Befundung.

		Krankheit		
		vorhanden	nicht vorhanden	
Test	positiv	richtig positiv a	falsch positiv b	a+b pos.prädiktiver Wert: $\frac{a}{a+b}$
	negativ	falsch negativ c	richtig negativ d	c+d neg.prädiktiver Wert: $\frac{d}{c+d}$
		a + c	b + d	n = a+b+c+d
Sensitivität:		Spezifität:		
$\frac{a}{a + c}$		$\frac{d}{b + d}$		

Abb.: 3.1 Konkordanz und Diskordanz von diagnostischem Test und Wirklichkeit

Der Quotient $a/(a + c)$ mißt den Anteil der Erkrankten, für die ein positives Testergebnis erhalten wurde und schätzt die Sensitivität des Tests.

Der Quotient $d/(b + d)$ beschreibt den Anteil der Gesunden, besser der nicht von der fraglichen Krankheit Betroffenen, für die der Test zum negativen Befund führte. Damit wird die Spezifität des Tests geschätzt.

Sensitivität und Spezifität sind die gebräuchlichsten prävalenz-unabhängigen Maßzahlen zur Beurteilung diagnostischer Tests. Prävalenzunabhängigkeit bedeutet, daß diese Werte im Idealfall nicht von der relativen Häufigkeit der Zielerkrankung abhängen. Zur Schätzung der Sensitivität wird lediglich eine Stichprobe Erkrankter benötigt, ebenso wird die Spezifität allein aus einer Stichprobe 'Gesunder' ermittelt. Tabelle 3.3 enthält die Zusammenstellung dieser und weiterer prävalenzunabhängiger Maßzahlen für die Bewertung diagnostischer Tests. In der Praxis kann die Annahme der Prävalenzunabhängigkeit von Sensitivität und Spezifität nicht in allen Fällen aufrechterhalten werden. Für die Anwendung diagnostischer Tests in der Nachsorge wird dies in Abschnitt 3.3.2 ausführlich diskutiert.

Tab. 3.3: Prävalenzunabhängige Indizes diagnostischer Tests, je eine Stichprobe Gesunder und Kranker erforderlich

Index	Schätzer
Sensitivität	$a/(a + c)$
Spezifität	$d/(b + d)$
Falsch-Negativ-Rate	$c/(a + c)$
Falsch-Positiv-Rate	$b/(b + d)$
YOUDEN-Index	Sensitivität + Spezifität - 1

3.1.2.2 Prävalenzabhängige Indizes

Im Abschnitt 3.1.1.1 wurde betont, daß sensitive Tests "im Prinzip" zum Ausschluß einer Erkrankung, spezifische Tests "im Prinzip" zur Bestätigung einer Verdachtsdiagnose geeignet seien. Die Einschränkung "im Prinzip" bezieht sich zum einen darauf, daß die Begriffe sensitiver bzw. spezifischer Test nicht exakt definiert sind. Zusätzlich aber, und das ist die entscheidende Randbedingung, wird die Wahrscheinlichkeit, daß eine Krankheit bei Vorliegen eines positiven Testergebnisses in Wirklichkeit vorhanden ist (bzw. beim negativen Testergebnis nicht vorhanden ist), durch die Häufigkeit der Krankheit in der Bezugspopulation, die Prävalenz, bestimmt.

In der Abbildung 3.1 kann die Prävalenz geschätzt werden als $\text{Prävalenz} = (a + c)/(a+b+c+d)$. Natürlich ist dieser Quotient nur

dann sinnvoll, wenn die Erkrankten und die Gesunden als Zufallsstichprobe aus einer gemeinsamen Grundgesamtheit gezogen wurden. Die wichtigsten prävalenzabhängigen Indizes diagnostischer Tests sind der Anteil der Erkrankten unter allen positiven Testausgängen, geschätzt als (positiver) prädiktiver Wert = $a/(a + b)$ und der Anteil der nicht Erkrankten an allen negativen Testergebnissen, geschätzt als negativer prädiktiver Wert = $d/(c + d)$.

Der positive prädiktive Wert beschreibt also die Wahrscheinlichkeit, daß bei positivem Testausgang die Krankheit tatsächlich vorliegt. Entsprechend gibt der negative prädiktive Wert die Wahrscheinlichkeit an, daß unter der gegebenen Prävalenz ein negativ befundeter Patient auch in Wirklichkeit krankheitsfrei ist bzw. höchstens an einer anderen Krankheit leidet. Diese und weitere prävalenzabhängige Indizes sind in Tabelle 3.4 zusammengestellt.

Tab. 3.4: Prävalenzabhängige Indizes diagnostischer Tests, repräsentative Stichprobe aus der Population erforderlich

Index	Schätzer
(positiver)prädiktiver Wert	$a/(a + b)$
negativer prädiktiver Wert	$d/(c + d)$
Falsch-Positiv-Rate, prävalenzabhängige Definition	$b/(b + d)$
Falsch-Negativ-Rate, prävalenzabhängige Definition	$c/(a + c)$
Genauigkeit	$(a + d)/(a + b + c + d)$
Prävalenz	$(a + c)/(a + b + c + d)$

Der Einfluß der Prävalenz auf den prädiktiven Wert von Tests ist in (24) ausführlich dargestellt. Abbildung 3.2 gibt hierzu ein Beispiel. Dargestellt wird die Anwendung eines Tests, etwa Röntgenuntersuchung der Lunge zur Auffindung von tumorösen Veränderungen in zwei unterschiedlichen Situationen, zuerst in der Normalbevölkerung und dann in einem Risiko- bzw. Verdachtskollektiv. Die Zahlenwerte für Sensitivität, Spezifität und Prävalenz wurden zu Demonstrationszwecken willkürlich gewählt.

Screening in der
Gesamtbevölkerung

Prävalenz: 0,5 %

Abklärung in einem
Verdachtskollektiv

Prävalenz 20 %

Test:

Sensitivität = Spezifität = 90 %

	krank	gesund	
Test pos.	45	995	1040
neg.	5	8955	8960
	50	9950	10000

	krank	gesund	
Test pos.	18	8	26
neg.	2	72	74
	20	80	100

Prädikt. Wert = $45/(45+995)=4\%$

Prädikt. Wert = $18/(18+8)=69\%$

Abb. 3.2: Prävalenzabhängigkeit des prädiktiven Wertes eines Tests.

Das Beispiel läßt erkennen, daß bei konstanter Sensitivität und Spezifität eines Tests durch Veränderung der Prävalenz unterschiedliche prädiktive Werte erzielt werden. Prävalenzunterschiede treten beispielsweise auch zwischen den unterschiedlichen Stufen des Versorgungssystems (Hausarzt, Facharzt, Klinik, Zentrum) auf.

3.1.2.3 ROC-Analyse

In der Realität wäre in der Situation der linken Seite der Abb. 3.2 zu überprüfen, ob und wie die Prävalenz angereichert werden kann, indem das Screening beispielsweise auf bestimmte Risikokollektive eingeschränkt wird. Vielfach ist jedoch die Anreicherung der Prävalenz des Zielereignisses nicht erreichbar. Veränderung der prädiktiven Werte des Tests sind dann nur durch Modifikation der prävalenzunabhängigen Indizes Sensitivität und Spezifität erreichbar. Da die meisten Testergebnisse, wie anfangs festgestellt, auf der Klassifikation primär nicht zweiwertiger Ergebnisse beruhen, ist eine Einflußnahme auf Sensitivität und Spezifität durch Veränderung der Klassifikation, z.B. des Schwellenwerts bei quantitativen Testergebnissen, in der Tat möglich. Allerdings geht die Verbesserung der Sensitivität eines Tests zu Lasten der Spezifität und umgekehrt, wie in Abb. 3.3 skizziert.

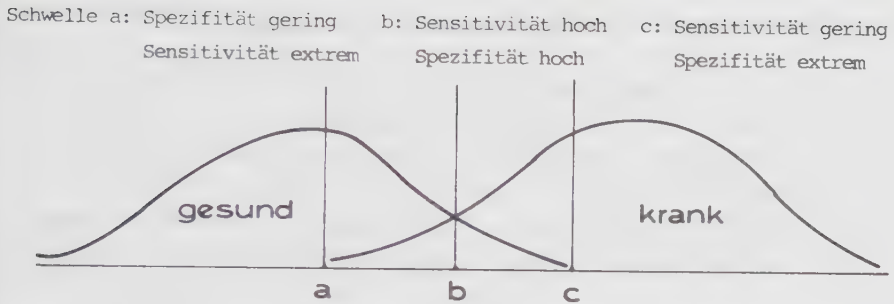


Abb. 3.3: Sensitivität und Spezifität in Abhängigkeit von der Wahl des Schwellenwerts für die Testbefundung (Klassifikators)

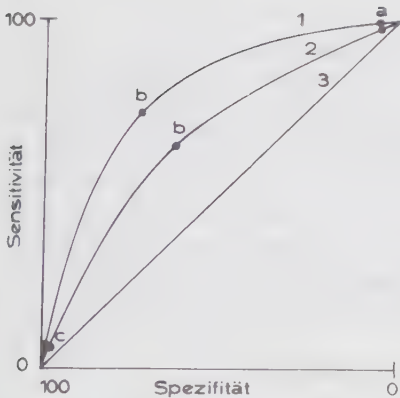


Abb. 3.4: ROC-Kurve zweier Tests. Test 1 ist überlegen im Vergleich zu Test 2. Die Arbeitspunkte a, b und c aus Abb. 3.3 sind schematisch dargestellt.

- Kurve 1: Test 1
- 2: Test 2
- 3: Zufallsentscheidung

Eine anschaulichere Form der Darstellung des Zusammenhangs zwischen Sensitivität und Spezifität eines Tests bildet die Receiver Operation Characteristic (ROC). Aus dieser Darstellung ist für jede gewünschte Sensitivität des Tests die erreichbare Spezifität ablesbar und umgekehrt, wie in Abbildung 3.4 dargestellt. Ein hypothetischer Test ohne falsch-negative und falsch-positive Ergebnisse würde zu einer Rechteckslinie führen. Je geringer die Trennschärfe, desto näher verläuft die ROC-Kurve an der Diagonalen, die das Ergebnis der Zufallsentscheidung darstellt. Im Vergleich zweier Tests ist daher der zur weiter links verlaufenden ROC-Kurve gehörige dem anderen überlegen.

3.1.2.4 Probleme der Testvalidierung

In der Praxis wird davon ausgegangen, daß Sensitivität und Spezifität Produkte eines Zufallsmechanismus sind, daß insbesondere jede Krankheitsausprägung an einem Patienten die gleiche a-priori-Wahrscheinlichkeit aufweist, durch den Test entdeckt zu werden. Jedoch ist die Voraussetzung, daß z.B. eine Sensitivität von 70% dadurch zustandekommt, daß jedes Vorhandensein der Zielerkrankung mit einer Wahrscheinlichkeit von 70% vom Test richtig erkannt wird, während sich andere Krankheitsmanifestationen der Erkennbarkeit systematisch entziehen, nicht zwingend. Einige frühe Arbeiten versuchten, die Sensitivität eines Tests als stetige Größe, die in Abhängigkeit vom Einzelfall Werte zwischen 0 und 100% annehmen kann, darzustellen (44, 73). Diese Modelle wurden später wegen der Komplexität des Ansatzes nicht weiterverfolgt, Sensitivität und Spezifität werden in der Regel als konstante Größen betrachtet (68).

Gegen dieses einfachere Modell bestehen keine Einwände, wenn die Parameter Sensitivität und Spezifität an einem für den geplanten Anwendungsbereich des Tests repräsentativen Kollektiv ermittelt wurden. Eine Mißachtung dieses Grundsatzes hat zur Folge, daß in den ersten Veröffentlichungen über ein neues diagnostisches Verfahren dessen Präzision überschätzt wird. Wenn derselbe Test später unter vielfältigen Bedingungen und unabhängig vom Urheber geprüft und angewandt wird, pendelt sich die Einschätzung von Sensitivität und Spezifität allmählich auf einem realistischen Niveau ein.

Für den Tumormarker CEA (karzinoembryonales Antigen) wurde dieser Prozeß auf Grund einer Sekundäranalyse der einschlägigen Literatur exemplarisch nachgezeichnet (49). Gesichtet wurde die englischsprachige Fachliteratur aus den Jahren 1969 bis 1973. Dies entspricht der Periode, in der mit dem Test weltweit Erfahrung gesammelt wurde. Berücksichtigt wurden alle Veröffentlichungen, in denen über mindestens 10 Patienten berichtet und der Test hinsichtlich seiner Wertigkeit zur Diagnosestellung eines Kolonkarzinoms untersucht wurde.

Auswertbar waren 17 Publikationen. Nur drei Arbeitsgruppen fanden eine hohe Sensitivität des Tests, 14 dagegen konnten dies nicht bestätigen. Über die Verteilung der möglicherweise für die Diagnosefindung relevanten Krankheitsparameter, z.B. TNM- bzw. Dukes-Stadium der untersuchten Patienten, wurde nur sporadisch berichtet. Insbesondere gaben die drei Autoren, die eine hohe Sensitivität des Markertests CEA fanden, keine Hinweise auf die Zusammensetzung des untersuchten Kollektivs.

Eine hohe Spezifität des Tests wurde nur in einer der 17 Veröffentlichungen gefunden. Auch in diesem Fall wurden keinerlei Hinweise über die Zusammensetzung des getesteten Kontrollkollektivs gegeben. Dagegen wird an anderer Stelle berichtet, daß beispielsweise für nicht an malignen Tumoren erkrankte Raucher in ca. 20% der Fälle mit positivem CEA-Test zu rechnen ist (23).

Die unzureichende Berücksichtigung der Stichprobenabhängigkeit von Sensivität und Spezifität eines Tests ist also verantwortlich für eine anfängliche Überschätzung beider Parameter. Zwar ist es u.U. zweckmäßig, die Sensivität eines Tests zuerst an eindeutig kranken und die Spezifität an eindeutig gesunden Personen zu bestimmen, um erste Eindrücke über die mögliche Leistungsfähigkeit eines neuen Tests zu gewinnen. Verfehlt ist allerdings der Rückschluß, die so ermittelten Indizes für 'die' Sensivität bzw. Spezifität des Tests zu halten. Vor allem Aussagen zur Spezifität erfordern ein breites Spektrum von Kontrolluntersuchungen. In der zitierten Studie wurde dies am Beispiel des Tumormarkers CEA belegt.

Für die Spezifität ist die Annahme konstanter a-priori-Wahrscheinlichkeiten problematischer als für die Sensitivität. Ergebnis breit angelegter Validierungsstudien könnte u.U. sein, daß die Spezifität eines Tests nur unter gewissen Voraussetzungen befriedigt, während bei Fehlen dieser Bedingungen die Anwendung des Tests nicht empfehlenswert ist. Ein Test kann für bestimmte differentialdiagnostische Probleme hervorragende Eigenschaften besitzen, für andere Anwendungen aber nahezu ungeeignet sein. Ebenso könnte die Spezifität gegen alle bis auf eine oder wenige Kontrollgruppen befriedigen. Solche Tests könnten zum Screening wertvoll, für Differentialdiagnostik dagegen ohne große Attrakti-

vität sein. Entscheidend ist, daß diese Effekte bekannt sind.

In der Krebsfrüherkennung kann aber auch die Annahme der konstanten Sensitivität Probleme aufwerfen. Für Tumoren bis zu einer bestimmten, vom Verfahren abhängigen Mindestgröße ist die Entdeckungswahrscheinlichkeit Null. Mit zunehmender Tumorgroße wächst dann auch die Wahrscheinlichkeit der Entdeckung, um irgendwann, wenn der Tumor eine hinreichende Größe erreicht hat, einen Wert nahe 100% zu erreichen. Bei 'der' Sensitivität eines Tests zur Tumorfriherkennung handelt es sich um einen von der Zusammensetzung des Kollektivs, d.h. z.B. von der Frequenz der Screens, abhängigen Durchschnittswert dieser individuellen Entdeckungswahrscheinlichkeiten.

Ein Beispiel für die stichprobenabhängige Variation der Sensitivität bildet der Test auf okkultes Blut im Stuhl. Zwar ist für ein Kollektiv diagnostisch gesicherter Dickdarmkarzinome die Sensitivität des Tests leicht zu ermitteln. Fraglich ist jedoch, ob ein so ermittelter Wert reproduziert wird, wenn der Test im Screening benutzt wird (31).

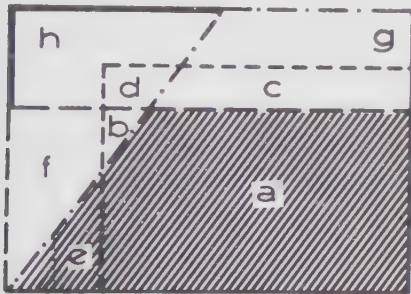
3.2 Präzision kombinierter Teststrategien in der Nachsorge

3.2.1 Test und Referenzkriterium

Um durch eine einzige Untersuchung zu einer eindeutigen Entscheidung in der Tumornachsorge, im Screening oder überall dort zu kommen, wo Gesunde hinsichtlich des Vorhandenseins einer bestimmten Veränderung untersucht werden, wären Tests erforderlich, die sowohl 'sensitiv' als auch 'spezifisch' sind. Wenn für eine Problemstellung überhaupt ein solcher Test verfügbar ist, so ist dessen Durchführung meist mit erheblichen Belastungen oder Risiken für den Patienten verbunden. Denn je präziser die Wirklichkeit durch einen Test abgebildet werden soll, um so schwerwiegender ist häufig der dazu notwendige diagnostische Eingriff. Gerade bei der Untersuchung Gesunder sind für die Anwendung solcher Tests Grenzen gesetzt. Anstelle der sofortigen Durchführung eines diagnostisch optimalen, aber risikoreicheren, belastenderen oder auch teureren Tests werden weniger anspruchsvolle Untersuchungen bevorzugt.

Ein akzeptabler Test mit hoher Sensitivität wird primär an allen zu untersuchenden Personen durchgeführt. Erbringt dieser ein negatives Ergebnis, so gilt die gesuchte Veränderung als ausgeschlossen. Bei positivem Testausgang gilt der nächste diagnostische Schritt dem Ausschluß falsch positiver Befunde, denn ein positives Ergebnis im Screening läßt allein nicht auf das Vorhandensein des Zielereignisses schließen (vgl. Abb. 3.2). Daher wird der Verdacht durch ein Referenz-Kriterium mit hoher Spezifität abgeklärt.

Wenn hierbei die Sensitivität dieser Referenzmethode der Sensitivität des Ausgangstests unterlegen ist, werden im Zuge der Abklärung nicht nur falsch positive, sondern auch zahlreiche richtig positive Ergebnisse eliminiert.



- · — · — Referenztest pos.
- — — — — Screeningtest pos.
- - - - - Zielereignis

Abb. 3.5: Positivbereich bei Kombination von Screening- und Referenztest. Wegen der sequentiellen Strategie werden die zum negativen Screeningtest gehörigen Zellen c, d, g und h nicht explizit erzeugt.

Unter der Annahme einer eindeutig definierten Krankheit ergeben sich die in Abb. 3.5 gezeigten Fälle. Als positiv werden alle im Screening positiven Befunde gewertet, die durch das Referenzkriterium bestätigt wurden (Abb. 3.6).

		Referenztest			
		RP	FP	RN	FN
Screening-Test	RP	a			b
	FP		e	f	
	RN		(g)	(h)	
	FN	(c)			(d)

Abb. 3.6: Kombination von Ergebnissen des Screening-Tests mit den Resultaten des Referenztests (vgl. Abb. 3.5)

Richtig positiv sind bei dieser Festlegung des positiven Bereichs die mit a, falsch positiv die mit e gekennzeichneten Fälle. Die Fälle b, c und d sind aus unterschiedlichen Gründen falsch negativ, während durch f, g und h richtig negative Befunde gekennzeichnet sind.

Sensitivität und Spezifität des Screening-Tests lauten:

$$\text{Sens}_T = (a+b)/(a+b+c+d)$$

$$\text{Spez}_T = (h+g)/(e+f+g+h)$$

Die entsprechenden Indizes der Referenzmethode lauten:

$$\text{Sens}_R = (a+c)/(a+b+c+d)$$

$$\text{Spez}_R = (h+f)/(e+f+g+h)$$

Für die Kombination bei sequentieller Anwendung ergibt sich:

$$\text{Sens}_K = a/(a+b+c+d)$$

$$\text{Spez}_K = (f+g+h)/(e+f+g+h)$$

Die Kombination eines Screening-Tests mit einer anschließenden Referenzmethode bewirkt also eine Erhöhung der Spezifität, die u.U. zu Lasten der Sensitivität geht. Es ist deshalb erforderlich, daß nicht nur die Spezifität, sondern auch die Sensitivität der Referenzmethode hoch ist, wenn der Anteil b klein gehalten werden soll. Ein negatives Testergebnis des Screening-Tests führt bei dieser Strategie nicht zur weiteren Abklärung, so daß die Fälle c, d, g, h in der unteren Hälfte von Abb. 3.6 in Wirklichkeit nicht erzeugt werden. Zur Beschreibung des Einflusses dieser Teststrategie auf Sensitivität und Spezifität ist dies jedoch belanglos.

3.2.2 Parallele Anwendung mehrerer Tests

Als weitere Anwendung kombinierter Tests ist die Durchführung mehrerer Untersuchungen zu nennen. Dies kann einerseits zur Erhöhung der Sensitivität dienen. Das Gesamtergebnis wird in diesem Fall als positiv angesehen, wenn einer der eingesetzten Tests ein positives Ergebnis zeigt (Abb. 3.7). Ein nachgeschalteter Referenztest könnte anschließend der Verifikation dieses Befundes dienen.

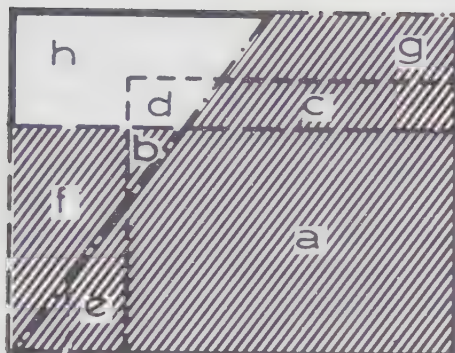


Abb. 3.7: Kombination zweier Tests zur Erhöhung der insgesamt erreichten Sensitivität zu Lasten der Spezifität

Die Kombinationsmöglichkeiten für die mit beiden Tests ermittelten Ergebnisse folgen dem Schema der Abb. 3.6, wobei in diesem Fall sämtliche Kombinationen a bis h tatsächlich erzeugt werden. Das Zielereignis wurde im Einklang mit der Absicht, die Sensitivität der Kombination zu maximieren, als Vereinigung der Positiv-Bereiche beider Tests definiert. Als positiv werden also die Fälle a, b, c, e, f und g gefunden. Die Sensitivität dieser Kombination beträgt

$$\text{Sens}_K = (a+b+c)/(a+b+c+d), \text{ die Spezifität nur noch} \\ \text{Spez}_K = h/(e+f+g+h).$$

Welche Größenordnung der durch Kombination von Untersuchungen erhaltene Zugewinn an Sensitivität hat, ist nicht allgemein zu beantworten. Die Annahme, daß die benutzten Tests statistisch unabhängig voneinander operieren, ist unwahrscheinlich. Ebenso unwahrscheinlich ist ein Null-Gewinn, der bei $b = 0$ bzw. $c = 0$ gegeben wäre. Realistisch für die Abschätzung der kombinierten Sensitivität ist die Beziehung

$$\max (\text{Sens}_1, \text{Sens}_2) < \text{Sens}_K < (\text{Sens}_1 + \text{Sens}_2 - \text{Sens}_1 \cdot \text{Sens}_2).$$

Dies beinhaltet, daß der erwartete Zugewinn an Sensitivität größer als Null, aber kleiner als unter der Annahme der Unabhängigkeit dieser testspezifischen Falsch-Negativ-Bereiche ist.

In (53) wird der empirisch ermittelte Zugewinn für eine konkrete Anwendung quantifiziert, nämlich die parallele Untersuchung der Leber durch Sonographie und Szintigraphie. Es wird eine Steigerung der "Treffsicherheit", definiert als "Summe der richtig positiven und richtig negativen Ergebnisse, normiert auf die Gesamtzahl der Untersuchungen", von 89% (Sonographie) und 80% (Szintigraphie) auf zusammen über 90% angegeben. In der Kombination erreichen die Tests damit Ergebnisse, die denen der Computertomographie (93%) vergleichbar sind.

Unter anderen Voraussetzungen kann die Verwendung paralleler Tests dem Ausschluß von falsch positiven Befunden, also der Erhöhung der Spezifität dienen. Dieser Fall tritt z.B. ein, wenn die benutzten Untersuchungen hinsichtlich ihrer Präzision etwa gleichwertig sind und kein zusätzliches, eindeutig überlegenes Untersuchungsverfahren als Referenzmethode zur Verfügung steht.

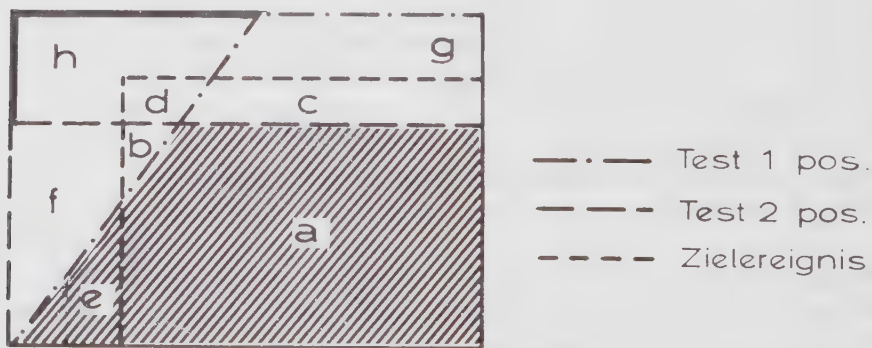


Abb. 3.8: Kombination zweier Tests zur Erhöhung der Spezifität zu Lasten der Sensitivität

Die Kombinationsmöglichkeiten von Befunden aus beiden Tests werden wiederum durch Abbildung 3.6 beschrieben. Als positiv gelten nun die Fälle a und e. Die Kriterien für positiven Gesamtbefund entsprechen also formal der Situation der Abb. 3.5. Praktisch unterscheidet sich das Vorgehen aber an zwei Stellen. Erstens werden die Tests parallel und nicht sequentiell durchgeführt. Würden die ungleichwertigen Tests in Abb. 3.5 simultan ausgeführt, so entspräche der positive Bereich der Gesamtbefundung dem positiven Bereich der Referenzmethode, und der Screeningtest wäre überflüs-

sig. Die Zellen c und g gehören in Abb. 3.6 nur deshalb zum Negativ-Bereich, weil sie wegen der sequentiellen Strategie nicht erzeugt, d.h. nicht von h und d unterschieden werden können.

Der zweite Unterschied zu Abb. 3.5 besteht im Ausmaß der Überlap-pung der Positiv-Bereiche beider Tests. Abb. 3.8 beschreibt den Fall, daß die Sensitivität der Gesamtbefundung erheblich hinter der Sensitivität der einzelnen Tests zurücksteht. Dafür ergeben sich für die Spezifität der Gesamtbefundung analoge Ergebnisse, wie in Abb. 3.7 für die Sensitivität. Wiederum gilt, daß die Wahrscheinlichkeiten für falsch-positive Befunde beider Tests in der Praxis nicht unbedingt als voneinander unabhängig angesehen werden dürfen.

3.2.3 Testwiederholung

Wichtiger Spezialfall kombinierter Teststrategien ist die Wieder-holung von Tests, die unmittelbar oder in geeignetem zeitlichen Abstand erfolgen kann. Ziel der Testwiederholung kann der Aus-schluß falsch-positiver (Believe-the-Negative Rule) oder der Ausschuß falsch-negativer Befunde (Believe-the-Positive Rule) sein (47). Im erstgenannten Fall stehen nur positive, im letzter-em nur negative Befunde zur Überprüfung durch Wiederholung an. Voraussetzungen und Gegenanzeigen solcher Wiederholungsuntersu-chungen sind in Abb. 3.9 zusammengestellt.

Als eigentliche Testwiederholung sind hierbei nur die Fälle im linken Teil der Abbildung zu bezeichnen. Wiederholte Testanwen-dungen im periodischen Screening und zur Beobachtung unklarer Be-funde wurden nur der Vollständigkeit halber aufgeführt.

Zur Abklärung fraglicher Befunde ist die Testwiederholung nur dann zweckmäßig, wenn die Fehlerursache nicht reproduzierbar an das Individuum gekoppelt ist. Dies wird in Abb. 3.9 als Validität bezeichnet. Dagegen können vor allem externe, mit der Reliabili-tät des Tests assoziierte Fehlerquellen, die z.B. mit Entnahme, Transport, Lagerung und und Analyse von Laborproben im Zusammen-hang stehen, durch Wiederholungsuntersuchungen angegangen werden.

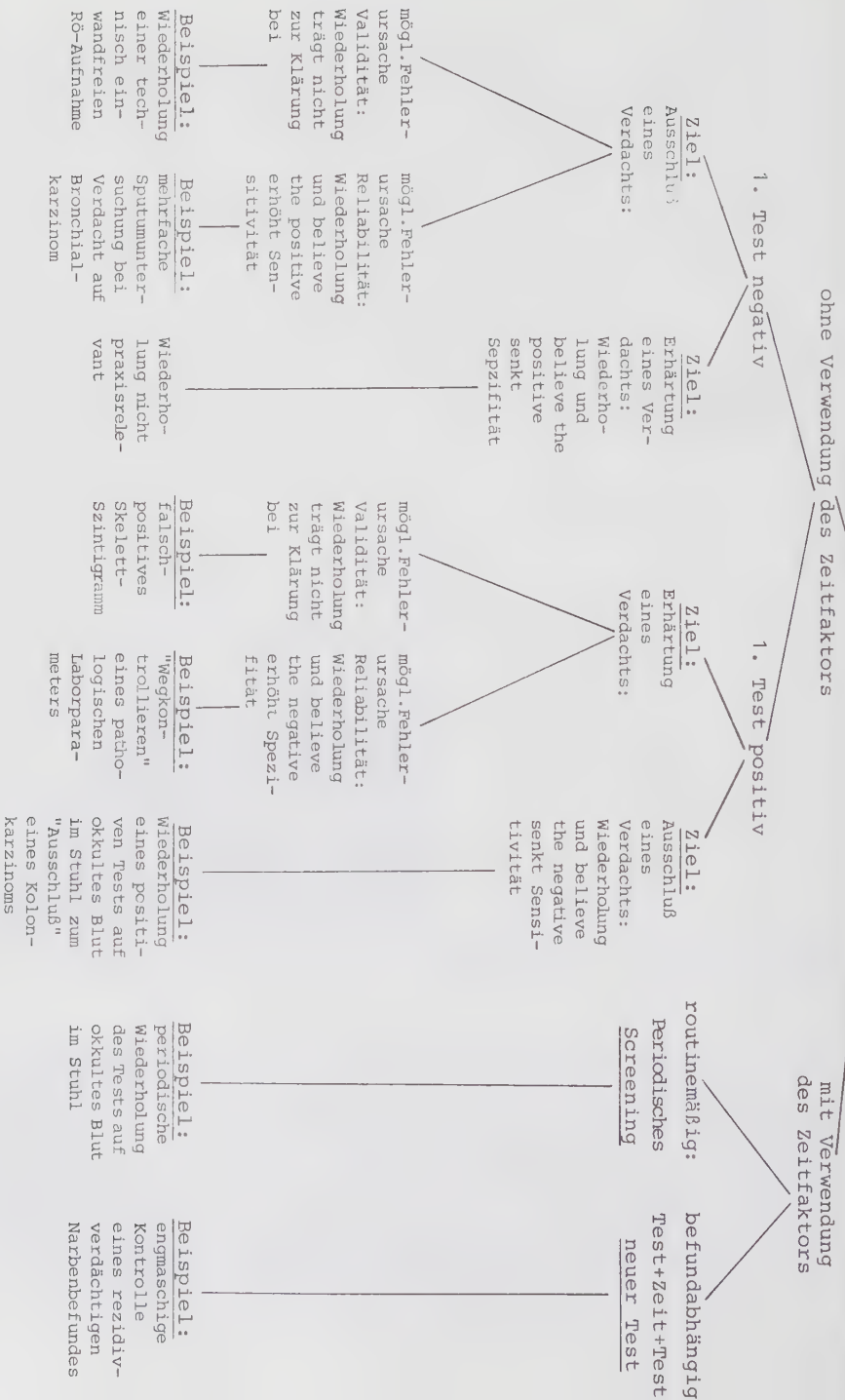


Abb. 3.9: Zur Logik von Testwiederholungen

Wird nach positivem Befund ein sensitiver, jedoch wenig spezifischer Test wiederholt, so schmälert dies die erwünschte Sensitivität, ohne daß ein relevanter Zugewinn an Spezifität erreicht wird. Der Test auf okkultes Blut im Stuhl ist ein bekanntes Beispiel für sinnwidrige Testwiederholung. Angemessen ist dieses Vorgehen bei primär negativem Ergebnis. Beispielsweise in der sputumzytologischen Diagnostik des Bronchialkarzinoms ist eine gegebenenfalls drei- bis fünfmalige Testwiederholung Teil der Routine.

Das Skelettszintigramm sei als Beispiel für validitätsbedingte Falsch-positiv-Befunde genannt. Häufig ist die differentialdiagnostisch korrekte Interpretation beobachteter Anreicherungen nicht oder nur mit Zusatzinformationen (z.B. aus Anamnese) möglich. Solche Unklarheiten könnten auch durch Wiederholung des Szintigramms nicht behoben werden.

Die in Abb. 3.9 als 'Wegkontrollieren' bezeichnete Wiederholung pathologischer Laborbefunde ist die möglicherweise häufigste Anwendung der Testwiederholung zur Erhöhung der Spezifität der Befundung. Dieses Problem hat im Zuge der Laborautomation durch Autoanalyser Technik an Bedeutung gewonnen(51).

3.3 Bestimmung der faktisch wirksamen Güteindizes

3.3.1 Empirische Ermittlung aus der Routine

Die Präzision eines diagnostischen Tests ändert sich mit Variation des Testumfeldes. Eine Bewertung der unter veränderten Bedingungen tatsächlich erzielten Trefferraten ist deshalb auch dann anzustreben, wenn überwiegend oder ausschließlich auf bekannte und anderweitig erprobte Tests zurückgegriffen wurde. Sofern keine ausreichenden Daten vorhanden sind, kann versucht werden, die Indizes Sensitivität und Spezifität aus Pilotstudien zu schätzen. Allerdings ist die interne Schätzung der Rate richtig positiver bzw. richtig negativer Fälle durch systematische Fehler beeinträchtigt, die nachfolgend beschrieben werden.

3.3.1.1 Wahrscheinlichkeit des richtig positiven Befundes

Voraussetzung für die Bewertung eines positiven Testergebnisses ist ein Außenkriterium. Wenn für den fraglichen Test eine Referenzmethode existiert, verursacht die Bewertung des positiven Tests keine grundsätzlichen Probleme: ein positives Testergebnis ist genau dann richtig positiv, wenn auch die daraufhin eingesetzte Referenzmethode ein positives Ergebnis nach sich zieht.

Hierbei muß berücksichtigt werden, daß auch 'Referenzmethoden' mehr oder weniger fehleranfällig sein können. Die Bewertung eines positiven Testergebnisses ist letztlich eine unter Berücksichtigung relevanter Begleitumstände getroffene subjektive Entscheidung. Die unverzerrte Beobachtung der Wirklichkeit ist nicht möglich. Im folgenden soll unter Referenzkriterium der auf den positiven Test folgende Aklärungsprozeß mit anschließender Entscheidung verstanden werden.

Mit der Notation

T = Test positiv

R = Referenzkriterium positiv

K = Krankheit

sowie

$\bar{K}$ = Krankheit nicht vorhanden

kann die Wahrscheinlichkeit $p(\bar{K}|T \wedge R)$ angegeben werden, daß ein bestätigtes positives Testergebnis in Wirklichkeit falsch positiv war. Diese Wahrscheinlichkeit hat nach der modifizierten Form des Bayes'schen Theorems den Wert

$$p(\bar{K}|T \wedge R) = \frac{p(R|\bar{K} \wedge T) \cdot p(\bar{K}|T)}{p(R|\bar{K} \wedge T) \cdot p(\bar{K}|T) + p(R|K \wedge T) \cdot p(K|T)}$$

Zur Vereinfachung wird angenommen, daß die Befundung des Referenzkriteriums R unabhängig von T ist, d.h.

$$p(R|\bar{K} \wedge T) = p(R|\bar{K}) \text{ und } p(R|K \wedge T) = p(R|K)$$

Damit ergibt sich

$$p(\bar{K}|T \wedge R) = \frac{p(R|\bar{K}) \cdot p(\bar{K}|T)}{p(R|\bar{K}) \cdot p(\bar{K}|T) + p(R|K) \cdot p(K|T)}$$

Unter realistischen Bedingungen erreicht der zu erwartende Befundungsfehler falsch-positiver Befunde eine beträchtliche Größenordnung. Gegeben seien beispielsweise ein Test T mit Sensitivität von 80% und Spezifität von 90% sowie ein Referenzkriterium mit Sensitivität und Spezifität von jeweils 98%. Beträgt die Prävalenz des Zielereignisses 5%, so liegt der prädiktive Wert des Tests T bei 30%, 70% der positiven Ergebnisse sind falsch-positiv. Der Befundungsfehler nach Anwendung der Referenzmethode erreicht damit 4,6%. Sinkt die Prävalenz auf 1%, entsprechend einem prädiktiven Wert von 7,5%, so erhöht sich dieser Fehler ceteris paribus auf 20%. Um den Befundungsfehler erneut unter 5% zu halten, muß nun die Spezifität des Referenzkriteriums über 99,5% liegen. Die Modellrechnungen in Abschnitt 6 werden belegen, daß niedrige Prävalenzen von 1-2% für konkrete Nachsorgestrategien realistisch sind.

Die Rate der vom Außenkriterium als positiv bestätigten Fälle ist daher prävalenzabhängig. Eine auf der Basis der gefundenen Fälle durchgeführte Schätzung von Sensitivität und Spezifität der benutzten Tests sowie der Prävalenz des Zielereignisses ist systematisch fehlerbehaftet.

3.3.1.2 Wahrscheinlichkeit des richtig negativen Befundes

Das negative Testergebnis eines Screeningtests wird nicht durch weitergehende Diagnostik überprüft. Eine Bewertung der erhaltenen negativen Testergebnisse durch indirekte Kriterien ist unter bestimmten Voraussetzungen möglich (38). Analog könnte eventuell im Zusammenhang mit Tumornachsorge verfahren werden.

Allerdings werden im Rahmen der Tumornachsorge teilweise auch negative Testergebnisse der weitergehenden Abklärung unterzogen, wenn ein anderweitig begründeter Verdacht auf Vorliegen des Ereignisses besteht. Dieser kann anamnestisch, durch unspezifische Untersuchungen oder anderweitige, positiv ausfallende Tests begründet sein. Der Anteil der überprüften negativen Tests ist daher keine Zufallsstichprobe aller negativen Tests. Falsch negative Tests sind aufgrund der Zusammensetzung dieser Stichprobe überrepräsentiert. Umgekehrt ist auch die verbleibende Menge der nicht überprüften negativen Tests nicht repräsentativ für die

Grundgesamtheit aller negativen Tests; da falsch negative Tests hier entsprechend unterrepräsentiert sind. Erst das gewogene Mittel der für beide Teilkollektive geschätzten Werte gibt die wahren Verhältnisse wieder.

Das Problem der Verzerrung des Außenkriteriums ist für Beurteilung von negativen Testergebnissen irrelevant. Die Wahrscheinlichkeit, daß das Zielereignis bei negativem Test und negativer Referenzmethode vorhanden ist, liegt unter 1%, wenn von realistischen Werten für Sensitivität und Spezifität für Test und Referenzkriterium ausgegangen wird.

Ein für die Beurteilung negativer Befunde typisches Problem dagegen ist die Dynamik des Krankheitsprozesses während der Abklärung eines Befundes. Setzt man für das Zielereignis eine Fünfjahresrate der Ereignisfreiheit von 60-80% an, so beträgt bei angenommener Exponentialverteilung die Wahrscheinlichkeit, daß das Ereignis innerhalb von 3 Monaten ab Durchführung des Tests neu auftritt, ca. 1 - 2,5%. Diese Zahl erscheint gering. Ihr steht jedoch ein prädiktiver Wert des negativen Tests gegenüber, der sich nur um die gleiche Größenordnung von 1 unterscheidet. Ein durch Anwendung der Referenzmethode nicht bestätigter negativer Test kann daher mit etwa gleicher Wahrscheinlichkeit durch ein falsch negatives Testergebnis oder durch Progredienz der Erkrankung nach richtig negativem Test verursacht sein.

3.3.2 Voraussetzungen und Probleme der Übertragbarkeit externer Schätzer

Sensitivität und Spezifität werfen nicht nur in der Phase nach Einführung eines Tests Fragen auf. Allgemein muß immer dann mit Veränderungen dieser Indizes gerechnet werden, wenn der Test in verändertem Umfeld benutzt wird. Neben der Zielpopulation gehören zum Testumfeld auch das ausführende Personal, gegebenenfalls die technische Ausstattung der ausführenden Stelle usw. sowie die begleitende Organisation. Im folgenden werden Effekte diskutiert, die für Nachsorgeprogramme typisch sind und Sensitivität und Spezifität von Tests verändern können. Solche Effekte sind bei der Planung von Nachsorgestrategien ins Kalkül zu ziehen.

Da praktische Erfahrungen mit Nachsorgestrategien bislang fehlen, können die aufgeführten Effekte (noch) nicht quantitativ mit Beispielen belegt werden. Naturgemäß ist auch nicht jeder dieser Effekte auf jeden Test anwendbar. Tests, deren Ergebnis von der Befundinterpretation durch den Untersucher abhängt, sind durch Umfeldveränderungen leichter zu beeinflussen als andere. Der größte Teil der Diagnostik im Rahmen der Nachsorge wird aber durch solche interpretationsabhängigen Tests abgedeckt.

3.3.2.1 Beeinträchtigung von Sensitivität und Spezifität durch Besonderheiten der Zielpopulation

Die Beschreibung von möglicherweise für die Wirkung eines Nachsorgeprogramms relevanten Faktoren beginnt mit testrelevanten Faktoren in der Zielpopulation eines Nachsorgeprogramms.

Tab. 3.5: Testrelevante Faktoren in der Zielpopulation, durch die Sensitivität und Spezifität in der Nachsorge beeinflusst werden können.

Faktor	Wirkung	Beispiel
1. Nachlassen der Aufmerksamkeit des Beobachters	Sensitivität ↓	Zytologiescreening
2. Veränderung des Schwellenwerts bei Ausbleiben erwarteter Treffer	Spezifität ↓	Rückgang der Prävalenz, z.B. durch Wechsel von der Klinik in die freie Praxis
3. Spektrum der Krankheitsausprägungen	Sensitivität ↓	Test auf okkultes Blut für Rezidivtumoren?
4. Spektrum möglicher konkurrierender Veränderungen	Spezifität ↓	Rezidiv oder Therapiefolge im Röntgen?
5. Tragweite des positiven Ergebnisses	Sensitivität ↓	Zurückhaltendere Befundung bei infauster Prognose
6. Tragweite eines falsch negativen Ergebnisses	Spezifität ↓	Besondere Vorsicht bei aggressiver, aber behandelbarer Krankheit

(1): Die Prävalenz des gesuchten Ereignisses kann Sensitivität und Spezifität von Tests auf verschiedenen Wegen beeinflussen. Bekannt ist die Tatsache, daß die Aufmerksamkeit

des Beobachters nachläßt, wenn sie nur selten durch Erfolge belohnt wird und sich steigert, wenn die Erfolge häufiger werden. Diesem Faktor sind Tests unterworfen, deren Sensitivität von der Aufmerksamkeit des Beobachters abhängt, z.B. die Reihenuntersuchung von Zytologiepräparaten. Ähnlich kann sich auch die Voreingenommenheit des Untersuchers über die vermeintliche Prävalenz auswirken, die der tatsächlichen Prävalenz nicht zu entsprechen braucht.

- (2): In der Tumornachsorge kann ein anderer prävalenzabhängiger Effekt größere Bedeutung erlangen: die Schwellenwertmodifikation aufgrund beobachteter Trefferquoten. Gemeint ist die Veränderung der Beurteilungskriterien des Beobachters, wenn die beobachteten Trefferraten nicht seiner (wodurch auch immer gerechtfertigten) Erwartung entsprechen.

Der Effekt entspricht dem z.B. in (52) beschriebenen Expectation Bias. Er tritt mit hoher Wahrscheinlichkeit auf, wenn ein Nachsorgeprogramm Diagnostikempfehlungen ausspricht, die nicht mit dem natürlichen Krankheitsverlauf, vor allem den realen Metastasierungsmustern, im Einklang stehen. Der durchführende Arzt, der die gemäß Programm erwarteten Beobachtungen niemals bestätigt findet, fühlt sich mit der Zeit verunsichert. Er könnte die Schuld bei sich selbst suchen, seinen Laborstandard, seine Fähigkeit als Befunder etc. anzweifeln und Schwellenwerte bzw. Befundungskriterien modifizieren. Wenn Nachsorge zusätzlich arbeitsteilig abläuft und er die anschließende Überprüfung der Verdachtsdiagnose nicht selbst durchführt, hängt es von der Kommunikationsstruktur zwischen den beteiligten Ärzten ab, ob wenigstens Informationen über die durch Modifikation hervorgerufene Falsch-positiv-Rate zum Untersucher zurückfließen. Auf die Kommunikation zwischen den Beteiligten wird weiter unten gesondert eingegangen.

- (3): Das Spektrum der Krankheitsausprägungen in der Nachsorge muß nicht mit dem Spektrum im Normalkollektiv des Tests übereinstimmen, sondern kann sich in der Verteilung der testrelevanten Kriterien von diesem unterscheiden. Wenn die Ursachen für falsch negative Testergebnisse nicht bekannt sind, ist

auch der Einfluß von Populationsunterschieden auf die Sensitivität schwer prognostizierbar. Der Test auf okkultes Blut im Stuhl könnte auch hierfür als Beispiel dienen. Denn da die Sensitivität des Tests im Screening bisher nicht wirklich validiert ist (31), ist auch nicht bekannt, wie Rezidivtumoren des Dickdarms durch den Test erfaßt werden. In der Krebsvorsorge ist der Test auch bei mittlerer Sensitivität brauchbar, da die Durchführung nicht belastet und falsch negative Ergebnisse daher kaum schaden. Dies gilt aber nicht mehr, wenn etwaige Symptome wegen des negativen Tests mißachtet werden. Besonders in der Nachsorge, wo jederzeit mit einem Tumorrezidiv zu rechnen ist, kann daher eine hohe Falsch-Negativ-Rate nicht unbedenklich sein.

- (4): Das Spektrum konkurrierender Veränderungen kann besondere Anforderungen an die Spezifität von Tests in der Tumornachsorge stellen. Für bildgebende Verfahren genügt es z.B. in der Nachsorge häufig nicht, Abweichungen vom physiologischen Zustand in einem Organ richtig zu erkennen. Denn gerade dann kann die entscheidende Frage lauten, ob 'harmlose' Therapiefolgen anzunehmen sind oder ein Tumorrezidiv vermutet werden muß. Dieses differentialdiagnostische Problem stellt andere Anforderungen an die Spezifität eines Tests als z.B. die Beurteilung des Organs im Rahmen des Staging vor der Primärtherapie.
- (5): Ob die Tragweite eines positiven Befundes für den Patienten die Sensitivität des Tests beeinflusst, hängt von der Beobachterabhängigkeit des Tests und gegebenenfalls der Beziehung zwischen Beobachter und Patient ab. Je geringer der betreuende Arzt die Chance zur erfolgreichen Intervention beurteilt, desto zurückhaltender wird er den durchgeführten Test befunden.
- (6): Mit umgekehrtem Vorzeichen gilt dies auch für die Tragweite eines falsch negativen Befundes. Patienten, die mutmaßlich oder tatsächlich einer Gruppe mit hohem Risiko angehören, für die aber auch chancenreiche Therapien verfügbar sind, erleben wohl eine aggressivere, d.h. sensitivere Testbefundung zu Lasten der Spezifität.

3.3.2.2 Fehler bei der Testdurchführung und autonome Eigenschaften des Untersuchers

Auch testbedingte Faktoren können Variation der Sensitivität verursachen. Ein diagnostischer Test ist keine deterministische Blackbox, die der Proband betritt, um sie mit einem Label 'positiv' oder 'nicht positiv' versehen wieder zu verlassen. Am Entstehen des Testergebnisses wirken heterogene Mechanismen mit, die, jeder für sich und unabhängig voneinander, mit Fehlern behaftet sein können. Tabelle 3.6 spezifiziert Effekte, die der Durchführung des Tests, der Person des Beobachters oder der Institution, an der die Untersuchung stattfindet, zuzuschreiben sind.

Tab. 3.6: Fehler bei der Testdurchführung und autonome Eigenschaften des Untersuchers, durch die Sensitivität und Spezifität in der Tumornachsorge beeinflusst werden können.

Faktor	Wirkung	Beispiel
1. Technischer Standard	Sensitivität ↓ und/oder Spezifität ↓	Veraltete Apparaturen
2. Normierung	Sensitivität ↓ und/oder Spezifität ↓	Fehlende Ringversuche
3. Risikobereitschaft	Spezifität ↓	Unterschiedliches Verhalten bei mehrdeutigem Untersuchungsergebnis
4. Fehlende Spezialisierung bzw. Übung	Sensitivität ↓ Spezifität ↓	Geringere Übung in der Interpretation bildgebender Tests bei Anwendung des Verfahrens in der Breitenversorgung
5. Falsche Indikationsstellung der Testanwendung	Sensitivität ↓	Wiederholung eines Guajak-Tests nach einmaligem positiven Ausgang mit falsch negativem Zweitergebnis
6. Unterkapazität	Sensitivität ↓	Zurückhaltendere Befundung, wenn Kapazität zur Befundabklärung fehlt
7. Überkapazität	Spezifität ↓	Aggressive Befundung, wenn Kapazitäten zur Abklärung unausgelastet sind

Eine Diskussion dieser Faktoren im Detail erübrigt sich, da es sich um bekannte Effekte handelt. Dennoch ist ihre Bedeutung für die Anwendung von Tests in der Breitenversorgung hoch. Dagegen findet in Validitätsstudien für diagnostische Tests günstigenfalls keiner dieser Mechanismen Eingang. Gute Validitätsstudien zur Ermittlung von Sensitivität und Spezifität eines Tests sind weder durch Kapazitätsprobleme belastet, noch werden andere als die jeweils optimalen Geräte oder Materialien verwendet. Der oder die Untersucher sind hinreichend qualifiziert, und auch die Beobachtervariabilität, die in solchen Studien erzeugt wird, entspricht nicht der Situation der Breitenversorgung. Vor unkritischer Übertragung publizierter Testindizes auf die Anwendung in Tumornachsorgeprogrammen muß daher gewarnt werden. Vergleichbare Probleme stellen sich hinsichtlich der Übertragbarkeit der Ergebnisse kontrollierter klinischer Studien, wobei die durch Breitenanwendung hervorgerufene Variabilität im Diagnostikbereich eher höher liegen dürfte als bei Verallgemeinerung von in randomisierten Therapiestudien geprüften Therapien.

3.3.2.3 Chancen zur Beeinflussung von Sensitivität und Spezifität durch geeignete Organisation der Nachsorgeprogramme

Untersuchungen, die in der Nachsorge durchgeführt werden, sind keine isolierten Ereignisse. Stellenwert und Aussagekraft der einzelnen Tests werden stets auch durch das Testumfeld bestimmt, in dem sie appliziert werden. Die Bezeichnung Nachsorgestrategie bringt die Interdependenz einzelner Untersuchungen zum Ausdruck. Die Organisation der Durchführung von Nachsorgeprogrammen entscheidet darüber, ob sich gewollte Interdependenzen zwischen Tests in der Praxis positiv auf Sensitivität und Spezifität der Untersuchungen auswirken können.

Tab. 3.7: Chancen zur Beeinflussung von Sensitivität und Spezifität durch geeignete Organisation der Nachsorgeprogramme

Faktor	Wirkung	Beispiel
1. Vorinformation	Sensitivität ↑	Vergleich der intraindividuellen Veränderung eines Parameters anstelle der Beurteilung einer Momentaufnahme
	Spezifität ↑	Vergleich eines fraglich positiven Befundes (z.B. Skelettszintigramm) mit einem früheren Referenzbefund
2. Testwiederholung im Verlauf der Nachsorge	Spezifität ↑	Vorsicht bei Tests, die schon einmal einen falsch positiven Befund verursachten
	Sensitivität ↑	Interpretation eines negativen Tumormarkerergebnisses nur dann, wenn die Testung vor Beginn der Primärtherapie ein pathologisches Resultat erbrachte
3. Parallele Tests	Sensitivität ↑	Lebersonogramm bei ansteigendem/konstantem CEA nach Dickdarmkarzinom
	Spezifität ↑	Lebersonogramm bei unauffälligem/auffälligem Lungen-/Skelett-/Lokalbefund nach Mammakarzinom
4. Möglichkeit zur weiteren Abklärung	Spezifität ↓	Engmaschigere Kontrolle wegen Entfernung des Wohnorts des Patienten nicht möglich, daher sofortige Abklärung fraglicher Befunde
	Sensitivität ↓	Vorsichtigere Befundung, wenn der Allgemeinzustand des Patienten eine weitere Abklärung positiver Ergebnisse nicht gestattet
5. Kommunikation mit Kollegen	Sensitivität ↑ Spezifität ↑	An anderer Stelle befindliche Vorinformation wird zugänglich
6. Übersichtlichkeit und Logik der Instruktionen	Sensitivität ↑ Spezifität ↑	Klare Instruktionen zur Durchführung der Nachsorge incl. weiterer Maßnahmen unterstützen die bestimmungsgemäße Anwendung des Programms
7. Qualitätskontrolle	Sensitivität ↑ Spezifität ↑	Reduktion von Abweichungen von der korrekten Applikation der Tests

(1): Die Wahrscheinlichkeit, einen Test richtig zu befunden, hängt in dreifacher Hinsicht vom Vorwissen ab. Qualifikation des Beobachters und Erwartung über die Häufigkeit des gesuchten Ereignisses wurden oben diskutiert. Hinzu kommt das Vorwissen über den bisherigen Krankheitsverlauf des Patienten. Dieses spielt nicht nur in der Befundung von Nachsorgeuntersuchungen eine Rolle. Beispielsweise wird jeder Frau empfohlen, etwa im Alter von 30 Jahren ein Basis-mammogramm als Referenz für spätere fragliche Befunde erstellen zu lassen. Analog wird in der Nachsorge häufig die Erstellung eines Szintigramms als Baseline-Untersuchung unmittelbar nach der Primärtherapie angestrebt. Solche Referenzbefunde können die Rate falsch positiver Befunde senken.

Typisch für die Nachsorge ist aber auch die engmaschige Anwendung desselben Tests innerhalb von Wochen oder Monaten. Je engmaschiger die zurückliegende Betreuung des Patienten durch den den Test durchführenden Arzt war, umso treffsicherer kann das Instrument des Tests in der Hand dieses Arztes werden.

An einem einfachen Beispiel soll dies erläutert werden. Betrachtet man "Sensitivität" und "Spezifität" des Karnofsky-Index für den Allgemeinzustand als Test für das Zielkriterium Überleben, so ist mit hoher Spezifität der Prognose zu rechnen. Ein Patient, dessen Allgemeinzustand tumorbedingt auf Werte um 30% oder darunter zurückfällt, hat kaum eine Chance, länger als Wochen oder Monate am Leben zu bleiben. Einleuchtend ist andererseits auch die Vermutung, daß sich nicht jeder Todesfall bereits Wochen vor seinem Eintritt durch Karnofsky-Werte um 30% ankündigt. Eine hohe Sensitivität dieses Parameters ist daher nicht zu erwarten. Genau dieses Ergebnis wurde in (72) anhand einer Querschnittserhebung des Karnofsky-Status und nachfolgender Beobachtung der Überlebenszeit bestätigt (Abb. 3.10).

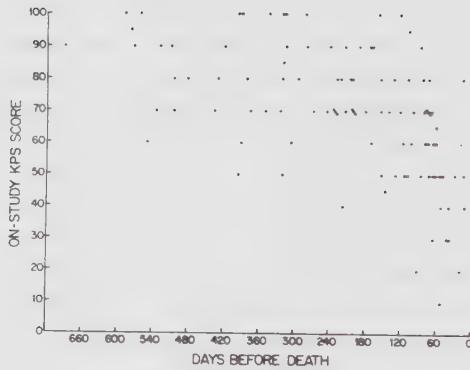


Abb. 3.10: Überlebenszeit des Patienten in Tagen in Abhängigkeit vom Karnofsky Index (entnommen aus (72); Definition s. Abb. 4.1)

Was besagt nun dieses Ergebnis für den Stellenwert der Anamneseerhebung im Rahmen regelmäßiger Nachsorgeuntersuchungen? Das Allgemeinbefinden eines Patienten gehört zu den Parametern, die vorwiegend durch Beobachtung ihres zeitlichen Verlaufs, d.h. ihrer Veränderungen seit dem letzten Arztbesuch, auf richtige Fährten lenken. Der plötzliche Übergang von 90 auf 70% kann prognostisch kritischer sein als ein konstanter Wert um 60%.

Dem Karnofsky-Index in der Nachsorge eine Sensitivität zuzuordnen zu wollen, ist daher vom Konzept her unsinnig. Sinnvoll wäre der Versuch, die Veränderung des Status als Test zu betrachten. Nur stellt sich dann die Frage der Formalisierbarkeit der Befundung. Vermutlich wird hierbei die Grenze zwischen Diagnostik und Institution erreicht oder überschritten.

- (2): Die Erfahrung mit früherer Anwendung des Tests am gleichen Patienten kann, ähnlich wie Baseline-Untersuchungen, vor falsch positiven Befunden schützen, z.B. in der Skelettszintigraphie. Ein neuerlicher fraglich positiver Befund kann gegebenenfalls mit dem bereits abgeklärten Befund verglichen werden, so daß eine erneute Falsch-positiv-Klassifikation

unterbleiben kann. Damit hat sich die Spezifität des Tests bei gleichgebliebener Sensitivität erhöht. Es ist jedoch nicht bei allen Tests entscheidbar, ob eine Wiederholung eines positiven Befunds dieselben Ursachen hat wie ein früherer falsch positiver Befund. Wenn dies nicht entscheidbar ist, geht die erneute Anwendung eines Tests nach falsch positivem Befund entweder zu Lasten der Sensitivität, nämlich wenn dem neuerlichen positiven Resultat schwerer geglaubt wird, oder zu Lasten der Spezifität, wenn der Test trotz erhöhter Wahrscheinlichkeit für neuerliche Falsch-positiv-Resultate unvoreingenommen weiter verwendet wird.

- (3): Die Gesamtbefundung parallel durchgeführter Tests wurde im Abschnitt 3.2 behandelt. Im Vorfeld und unabhängig vom Gesamtbefund wirkt sich die parallele Durchführung von Tests auf Sensitivität bzw. Spezifität der einzelnen Untersuchungen aus. Im Kontext einer Validierungsstudie wäre dieser Effekt als Bias zu klassifizieren. Er entspricht in etwa dem "Test Review Bias" nach (49), der entsteht, wenn die Interpretation des zu validierenden Tests durch die Kenntnis der Diagnose beeinflusst wird. Auf die Nachsorge bezogen ist lediglich der Einfluß der Diagnose durch den Einfluß der anderen, parallel anfallenden Testergebnisse ersetzt.

Anders als bei Validierungsstudien kann dieser Effekt in der Tumornachsorge durchaus erwünscht sein, indem verschiedene Tests synergistisch auf dasselbe Ergebnis hinweisen und gemeinsam zum positiven Befund ermutigen. Negative Auswirkungen ergeben sich dann, wenn der Stellenwert der parallel durchgeführten Tests, ihre 'Glaubwürdigkeit', asymmetrisch ist und dem präziseren Test weniger vertraut wird als dem unterlegenen.

Als praktisches Beispiel für die synergistische Verwendung paralleler Tests außerhalb der Nachsorge ist die histologische Befundung von Knochentumoren zu nennen. Auf dem Weg von der morphologischen Deskription zur Klassifikation eines Präparats, also zur histologischen Diagnose, ist das Röntgenbild des befallenen Knochens ein unverzichtbares Hilfsmittel. Morphologie und Röntgen gemeinsam sind zur Diagnose-

findung notwendig und hinreichend. Befundet man die beiden Tests aber gegeneinander blind, so ergibt sich ein unbefriedigendes Resultat.

- (4): Die Möglichkeit zur weiteren Abklärung verdächtiger Befunde ist ein weiterer Einflußfaktor auf Sensitivität und Spezifität. Für die Befundung eines Tests kann es erheblich sein, ob ein Patient bei negativem Ergebnis für engmaschige Kontrollen zur Verfügung steht, ob ihm bei positivem Ergebnis die nachfolgende Diagnosesicherung überhaupt zugemutet werden kann usw. Ein nahezu identischer Effekt wurde in Tabelle 3.5 als populationsabhängiger Faktor angegeben. Die nochmalige Nennung dieses Faktors in Tabelle 3.6 als organisationsbedingter Effekt ist berechtigt, wenn die Limitierung der weiteren Abklärung z.B. aufgrund des weiten Anfahrtsweges und des damit verbundenen Aufwandes für den Patienten entsteht. Nachsorge in Schwerpunktambulanzen führt teils zu Anfahrtswegen von über 200 km, während bei dezentraler und wohnortnaher Nachsorge diese Belastungen vermieden werden.
- (5): Auf einer anderen organisatorischen Ebene liegt die Kommunikation der in die Nachsorge involvierten Ärzte. Kommunikationsprobleme beginnen auf der untersten Stufe, der Kommunikation eines Arztes mit sich selbst via Krankenakte. Sie pflanzen sich fort in der Verständigung innerhalb eines Hauses, zwischen Kollegen und Abteilungen. Diagnostische Abteilungen bemängeln häufig, daß auf Anforderungsbelegen für diagnostische Leistungen Rubriken wie 'Fragestellung' von Station nicht oder nicht sorgfältig genug ausgefüllt werden. Die resultierenden Befunde erreichen dementsprechend nicht immer die Präzision, die unter Ausnützung aller vorhandenen Vorinformationen möglich gewesen wäre.

Kommunikationsbedarf in der Nachsorge entsteht vorwiegend im ambulanten Bereich. Erschwerend wirkt sich aus, daß die Art der notwendigen Kommunikation sich von dem Vorgehen unterscheidet, das bei der Lösung anderweitiger diagnostischer Probleme zum Ziel führt. Zwar werden auch außerhalb der Tumornachsorge Diagnosen häufig in Kooperation von zwei oder mehreren niedergelassenen Ärzten gestellt. Die Kommunikation

beschränkt sich jedoch in diesem Fall auf den einmaligen Austausch von Information. Ein Patient wird mit einer definierten Frage einem Kollegen überwiesen, dieser liefert den erbetenen Befund, und damit ist der Vorgang beendet.

Dagegen ist in der Nachsorge der Rückgriff auf frühere Befunde notwendig. Prospektiv ist nicht unbedingt absehbar, welche der momentan durchgeführten Untersuchungen später von Interesse sein werden. Unerwartet kann ein Arzt vor der Frage stehen, welche für ein bestimmtes Problem relevanten Vorbefunde für seinen Patienten vorliegen, in welchem zeitlichen Abstand sie erstellt wurden und wo sie gegebenenfalls verfügbar sind. Ein Nachsorgeprogramm, in dem dieser Rückgriff auf frühere Befunde, auch wenn sie bei Kollegen erstellt wurden, nicht reibungslos möglich ist, wird seiner Zielsetzung schwerer gerecht.

- (6): Auf welcher vielfältigen Weise Sensitivität und Spezifität durch geschickte Nachsorgestrategien beeinflussbar sind, wurde anhand der vorhergehenden Beispiele erläutert. Eine Kommunikationsstruktur, die nicht mit subtilen Details der Nachsorgestrategie Schritt hält, kann diese Effekte gefährden.

Vergleichbares gilt auch für die Information der beteiligten Ärzte über Inhalt und Hintergrund des durchzuführenden Programms. Trivial ist hierbei die Forderung, daß aus den Instruktionen klar erkennbar sein muß, wann welche Maßnahme durchzuführen ist. Wünschenswert, obwohl bisher kaum beachtet, sind darüberhinaus Hinweise auf zweckmäßige Abklärungsstrategien, auf differential-diagnostische Probleme und typische Symptome, verbunden jeweils mit Wahrscheinlichkeitsaussagen, soweit verfügbar. Derartige Hinweise fehlen meist. Eine Ausnahme bilden die in (67) publizierten Empfehlungen.

Auch durch kompakte Präsentation von Basisinformationen zu Diagnostik, Primärtherapie, Verlauf und Rezidivbehandlung des betreffenden Krankheitsbildes kann Problembewußtsein beim Nachsorgearzt erzeugt werden. Je transparenter die spezifischen Probleme dargestellt werden, um so leichter wird

die Logik einer Nachsorgestrategie nachvollzogen. Diesem Zweck u.a. dienen die Manuale zu Diagnostik, Therapie und Nachsorge, die in einigen Tumorzentren angeboten werden.

- (7): Dies sind Möglichkeiten zur begleitenden organisatorischen Unterstützung eines Nachsorgeprogramms. Sicherzustellen, daß der gewünschte hohe Standard der Durchführung tatsächlich erreicht wird, ist Aufgabe von Qualitätssicherungsmaßnahmen in der Nachsorge. Das Spektrum möglicher Maßnahmen ist reichhaltig, die in diesem Abschnitt diskutierten Aspekte zeigen Ansatzpunkte auf.

3.4 Zur pragmatischen Bedeutung von Sensitivität und Spezifität in der Nachsorge

Zu Beginn dieses Abschnitts wurde die Rolle von diagnostischen Tests in der Nachsorge diskutiert und dargelegt, welche Auswirkungen die häufigsten Diagnostikstrategien auf die Präzision der Befundung in der Nachsorge haben. Anschließend wurden Möglichkeiten und Probleme der Schätzung bzw. Prädiktion der Sensitivität eines Tests in der Nachsorge erörtert.

Hierbei beziehen sich Sensitivität und Spezifität auf das korrekte Erkennen des Vorliegens bzw. der Abwesenheit des Zielereignisses. Aus zwei benachbarten Blickrichtungen soll nun versucht werden, Kriterien anzugeben, unter welchen Bedingungen das (formale) Zielereignis, an dem Sensitivität und Spezifität des Tests gemessen werden, in der Nachsorge auch pragmatisch von Bedeutung ist. Hierzu wird zuerst untersucht, unter welchen Voraussetzungen das diagnostische 'Ziel'-Ereignis tatsächlich zu dem Ziel führt, eine Behandlung einleiten zu können. Anschließend werden Grenzen aufgezeigt, jenseits derer die Modellvorstellung eines eindeutig definierten Zielereignisses sich als problematisch erweist.

Ziel der folgenden Ausführungen ist es, auf (prävalenzunabhängige) Limitierungen der Aussagekraft von Sensitivität und Spezifität in der Nachsorge hinzuweisen. Insbesondere soll gezeigt werden, daß günstige Werte der genannten Indizes zwar notwendig, nicht aber in allen Fällen hinreichend sind, um den betreffenden Test als geeignet für die Tumornachsorge einzustufen zu können.

3.4.1 Diagnostische Information als Grundlage der Therapieplanung

Der Rückschluß auf das Vorliegen einer Krankheit erfolgt durch das Erkennen von Produkten und/oder Manifestationen dieser Krankheit. Häufig ist eine Trennung von Krankheit, Krankheitsmanifestationen und Krankheitsprodukten bedeutungslos. Für generalisierte Tumorerkrankungen dagegen ist diese Unterscheidung zweckmäßig. Unter 'Krankheit' soll im folgenden verstanden werden, daß sich an einem beliebigen, möglicherweise unbekannten Ort im Körper aktive Tumorzellen befinden. 'Krankheitsmanifestation' soll die Existenz von Tumorzellen an einem bekannten Ort bezeichnen. 'Krankheitsprodukt' ist als Sammelbezeichnung für alle Phänomene gewählt, die Rückschlüsse auf das Vorliegen der Krankheit bzw. einer Krankheitsmanifestation zulassen.

Test	⇒ Krankheits- produkt	⇒ Krankheits- manifestation	⇒ Krankheit
A: Urogramm	Durchfluß- hemmung des Ureters	intraabd. Tumor	Progression eines Hoden- tumors
B: AFP-Test	AFP ↑	?	Progression eines Hoden- tumors
C: Biopsie	-	Lokalrezidiv	Progression eines Mammaca
D: (Inspektion)	Atemnot	Lokalrezidiv	Progression eines Larynxca
E: (Anamnese)	Verfall des Patienten	?	Progression eines Bronchialca
F: (Inspektion)	-	Charakteri- stischer Tumorherd	Progression eines mal. Melanoms

Abb. 3.11: Wege zur Entdeckung einer Tumorprogression in der Nachsorge

⇒ Richtung des logischen Schließens

In der Tumornachsorge lautet die Zielkrankheit "Progression des Grundleidens". Krankheitsmanifestationen sind Ansammlungen von Tumorzellen in bestimmten Organen oder Strukturen. Krankheitsprodukte im oben definierten Sinne sind alle Folgeerscheinungen dieser Tumorzellen, z.B. biochemische Produkte oder mechanische Beeinträchtigungen. Einige Beispiele für den Prozeß der Diagnostizierung einer Tumorprogression sind in Abb. 3.11 zusammengestellt.

Im Fall A erkennt der diagnostische Test das Vorliegen einer Raumforderung, die unter Einsatz des Vorwissens als Tumorrezidiv gedeutet wird, welches - wiederum unter Verwendung des Vorwissens - als Progression eines bestehenden Hodentumors erkannt wird. Fall B führt am Ende zur gleichen Schlußfolgerung, ohne daß eine konkrete Krankheitsmanifestation erkannt wurde. Wo der Tumor lokalisiert ist, bleibt daher offen. Fall C beschreibt die direkte Beurteilung der Krankheitsmanifestation bei Fehlen von Krankheitsprodukten. Unter D, E und F schließlich wiederholen sich diese Wege, wobei der diagnostische Test eigentlich nicht mehr als solcher zu bezeichnen ist.

Die Schlußrichtung des Erkennens läuft vom Produkt über die Manifestation zur Diagnose der Krankheit. Entgegengesetzt verläuft die Krankheitsdynamik. Die 'Krankheit' erzeugt Krankheitsmanifestationen, die sich durch ihre Produkte im Körper bemerkbar machen. In der Mehrzahl der Fälle sind es erst die Krankheitsprodukte, die den Patienten beeinträchtigen und letztlich den schicksalhaften Ausgang der Krankheit hervorrufen.

An der Krankheitsdynamik orientiert sich die Therapie. Die Behandlung der Krankheitsprodukte hat symptomatischen Charakter. Kausale Therapie richtet sich gegen die Krankheit insgesamt oder zumindest einzelne Manifestationen. Wird die Krankheit insgesamt beseitigt, so kann von kurativer Therapie gesprochen werden, andere Maßnahmen werden als palliativ klassifiziert. In der Praxis ist diese Abgrenzung allerdings u.U. schwierig. Häufig kann die benötigte Information nicht gewonnen werden, so daß nicht entscheidbar ist, ob Tumorzellen im Körper verblieben sind oder nicht.

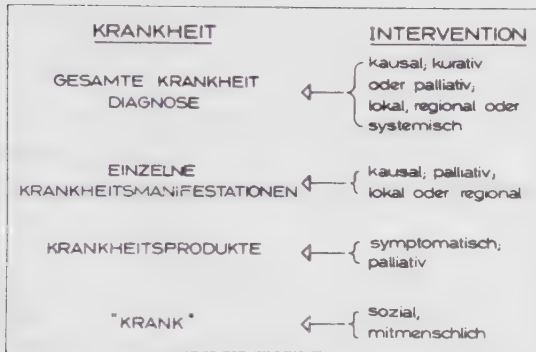


Abb. 3.12: Interventionsebenen bei Vorliegen einer Tumorprogression: Liefert der diagnostische Test jeweils die adäquate Information, um auf der nach dem Stand des medizinischen Wissens optimalen Ebene agieren zu können?

Ziel der Diagnostik in der Nachsorge ist die Bereitstellung der zur Indikationsstellung und Bewertung von Therapiemaßnahmen erforderlichen Fakten. Aus der Gegenüberstellung der Abbildungen 3.11 und 3.12 geht jedoch hervor, daß der Weg, über den ein diagnostischer Test angreift, nicht immer zu dem Punkt führt, an dem auch die verfügbare Behandlung ansetzen könnte. Unterstellt man die Richtigkeit eines Testergebnisses, so ist für dessen Handlungsrelevanz entscheidend, daß die lokale Unschärfe des Befundes nicht größer ist als der durch den erforderlichen bzw. möglichen Eingriff erfaßte Anteil des Organismus. Ein richtig positives Ergebnis bei der alkalischen Phosphatase als Hinweis auf Skelettmetastasen wäre z.B. (theoretisch) hinreichend für eine systemische Therapie beim Mammakarzinom, nicht jedoch zur Indikationsstellung einer palliativen Lokalbestrahlung, für die die genaue Lage des Metastasenherdes und eventuell Information über Bedrohlichkeit und Aggressivität des Tumors erforderlich sind.

Daraus ist abzuleiten, daß die pragmatische Relevanz eines Testergebnisses einen von Sensitivität und Spezifität unabhängigen Faktor der Bewertung darstellt. Wenn die durch einen Test bereitgestellte Information nicht zur Einleitung sinnvoller Maßnahmen genügt, so ist dieser Befund - selbst bei einer gedachten Spezifität von 100% - für das augenblickliche Handeln bedeutungslos.

3.4.2 Diagnostisch krank, therapeutisch gesund?

Bisher wurde vorausgesetzt, daß für jeden Patienten eindeutig feststehe, ob er an einer Krankheit leidet oder nicht. Beispielsweise bei der Erkennung von Infektionskrankheiten ist diese Annahme zweckmäßig. Degegen kann sie mit dem Grundlagenwissen über Entstehung und Verlauf von Tumorkrankheiten nur schwer in Einklang gebracht werden.

Ob das Modell in der Tumornachsorge pragmatisch zweckmäßig ist oder nicht, hängt von der konkreten Situation ab. Zweckmäßig ist die Annahme, wenn Test und Referenzmethode an Hand desselben Krankheitsparameters entscheiden, wie z.B. bei der Kombination eines Tests auf okkultes Blut im Stuhl mit der Colonoskopie als Referenzmethode. Wenn Tumorzellen ursächlich für eine Blutbeimengung sind, so muß ein Tumor von einer gewissen minimalen Größe vorliegen, der auch durch Endoskopie überprüfbar ist. Auch die umgekehrte Schlußrichtung ist sinnvoll, die Frage, wieviel Prozent der endoskopisch sichtbaren Tumoren bzw. Tumorrezidive sich durch okkultes Blut im Stuhl zu erkennen geben.

Weniger zweckmäßig wäre die Annahme einer eindeutig definierten Zielerkrankung zur Bewertung der Kombination des Tumormarkers CEA und der Colonoskopie als Referenz. Erstens beschränkt sich die diagnostische Aussagekraft dieses Tumormarkers nicht auf Tumormanifestationen im Dickdarm, so daß Colonoskopie allein als Referenzmethode ohnehin unzureichend wäre.

Zweitens steht ein anderer Gesichtspunkt der Kombination von CEA als Screeningtest und Endoskopie als Referenzmethode im Wege. Denn Ziel der Entwicklung und Anwendung von Tumormarkern ist gerade die Erschließung neuer diagnostischer Dimensionen, die Entdeckung von Tumoren, die wegen ihrer minimalen Größe durch herkömmliche Verfahren nicht verifizierbar sind. Mit der Anwendung des Tumormarkers CEA zur Verlaufsüberwachung verbindet sich daher eine Änderung des Krankheitsbegriffs, eine Erweiterung des Spektrums der Zielereignisse.

Zwar entspricht das diagnostische Vorgehen dem in Abb. 3.5 geschilderten Procedere. Zuerst wird der Markertest durchgeführt,

dessen Ergebnis über die Durchführung der Endoskopie als weiterführender Diagnostik entscheidet. Geändert hat sich jedoch die Interpretation der möglichen Untersuchungsergebnisse.

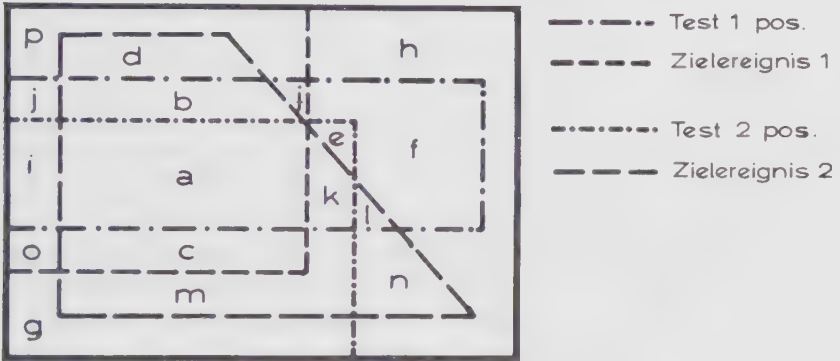


Abb. 3.13: Anwendung zweier Tests mit unterschiedlichen test-spezifischen Zielereignissen

Wegen der nicht identischen Zielereignisse für die Tests ergibt sich nun eine 4x4-Feldertafel (Abb. 3.14). Kombinationen wie richtig positiv/richtig negativ usw. sind nun durch die nur noch partielle Überlappung der Zielereignisse möglich geworden.

		Test 2, z.B. Endoskopie			
		RP	FP	RN	FN
Test 1, z.B. CEA	RP	a	i	j	b
	FP	k	e	f	l
	RN	m	g	h	n
	FN	c	o	p	d

Abb. 3.14: Mögliche Gesamtergebnisse bei simultaner Verwendung von Tests gemäß Abb. 3.13

Eine Berechnung der 16 Zeilen a, b, ..., p ist nicht ohne zusätzliche Annahmen über die Interdependenz der testspezifischen

Zielereignisse durchführbar. Eine Vereinfachung des Schemas ergäbe sich, wenn eines der Zielereignisse eine echte Teilmenge des anderen wäre, also z.B. für alle endoskopisch sichtbaren Tumoren auch eine Erhöhung des CEA anzunehmen wäre. In diesem Fall würden beispielsweise die Zellen k, l, m und n verschwinden. Gerade davon aber kann im gewählten Beispiel und in der Mehrzahl vergleichbarer Fälle nicht ausgegangen werden.

Am Beispiel der Tumormarker stellt sich die Frage, ob Sensitivität und Spezifität in allen Fällen der geeignete Rahmen sind, um die Arbeitsweise und Wirkung diagnostischer Maßnahmen in der Nachsorge zu beschreiben. Die Aufgabe eines Tumormarkers besteht darin, Krankheitsprozesse aufzuspüren, die durch andere Methoden nicht verifizierbar sind. Einem Tumormarker eine Sensitivität zuzuordnen, setzt daher zweierlei voraus.

Erstens ist ein zeitlicher Rahmen zu definieren, innerhalb dessen das vom Tumormarker vorausgesagte Auftreten der Erkrankung mit Hilfe anderer Verfahren verifizierbar sein muß. Damit wäre eine prognostische Fähigkeit des Tests meßbar.

Eine Gleichsetzung dieser prognostischen Leistung mit diagnostischer Treffsicherheit würde zweitens weitere Annahmen erfordern. Es müßte angenommen werden, daß alle richtig erkannten mikroskopischen Tumorherde innerhalb dieses Zeitraumes tatsächlich zu klinisch faßbaren Tumormanifestationen heranwachsen und zugleich alle zu einem bestimmten Zeitpunkt auftretenden Manifestationen vorher hätten entdeckt werden können. Die beiden Zielereignisse der Abb. 3.13 müßten also durch zeitlich versetzte Beobachtung quasi zur Deckung gebracht werden können. Das hierzu erforderliche Grundlagenwissen ist nicht vorhanden. Je weiter das Testergebnis vom klinisch faßbaren entfernt liegt, um so schwieriger ist die quantitative Beschreibung der diagnostischen Fähigkeiten des Tests.

Sind Tumormarker und andere Frühwarnsysteme, z.B. auch die sorgfältige Erhebung prognostisch relevanter Faktoren, daher ohne Wert? Die Antwort hängt u.a. davon ab, ob und in welchen Phasen der Krankheitsverlauf nach der Primärtherapie überhaupt wirkungsvoll beeinflussbar ist. Festzustellen ist, daß Sensitivität und

Spezifität allein nicht für alle diagnostischen Maßnahmen im Rahmen der Nachsorge als Beurteilungskriterium ausreichen.

Ein Gedankenexperiment mag diese Überlegung abrunden: Wie wäre ein Test zu beurteilen, der auf das Vorhandensein von nur wenigen Tumorzellen im Körper anspricht?

Die Antwort fällt leichter als erwartet, denn solche 'Tests' existieren bereits und zwar auf der Basis der statistischen Beobachtung der Überlebenszeit, z.B. nach Mammakarzinom. Beispielsweise ist es für Patienten mit mehr als drei positiven axillären Lymphknoten bei Diagnosestellung als nahezu sicher anzusehen, daß nach der Primärtherapie Tumorzellen im Körper verblieben sind. Der prädiktive Wert dieser a-priori-Vermutung beträgt nahezu 100%. Dennoch hat diese Aussage bisher nicht zu allgemein anerkannten Therapien geführt, durch die sich die Prognose des axillär positiven Mammakarzinoms verbessert. Die anhaltende Diskussion über den Stellenwert der adjuvanten Chemotherapie des Mammakarzinoms verdeutlicht dies. Dies ist ein bekanntes Beispiel für die in der Überschrift gestellte Frage.

3.5 Prävalenz latenter Tumorprogression in der Nachsorge

Der Vorhersagewert eines Tests ist abhängig von der Prävalenz des Zielereignisses. Für die Planung eines Nachsorgeprogramms ist die adäquate Einschätzung der Prävalenz daher von besonderer Wichtigkeit. Je höher die Prävalenz der Zielerkrankung liegt, um so effizienter arbeiten die eingesetzten Tests.

Zwischen Inzidenz, Prävalenz und Dauer einer Erkrankung besteht der Zusammenhang $\text{Prävalenz} = \text{Inzidenz} \times \text{durchschnittliche Dauer}$. 'Erkrankung' im Sinne der Tumornachsorge ist die präklinische, d.h. asymptomatische, jedoch durch Tests diagnostizierbare Phase der Generalisierung der Tumorerkrankung. Schwierigkeiten der Definition bzw. Abgrenzung von Zielereignissen und damit Unschärfen des Krankheitsbegriffs wurden oben diskutiert. Diese Phase endet durch Früherkennung oder klinische Manifestation einer Progression bzw. durch den Tod des Patienten.

Die Datenlage zur Vorhersage der zu erwartenden Prävalenz an Fernmetastasen bzw. lokoregionären Rezidiven in einem Nachsorgeprogramm ist erfahrungsgemäß schlecht. Dennoch und gerade deshalb ist die Erschließung und Nutzung vorhandener Datenquellen zur Verlaufsbeschreibung von Tumorerkrankungen von besonderer Bedeutung für die Planung von Nachsorgeprogrammen.

Aus diesem Grund werden Nutzungsmöglichkeiten der verfügbaren Daten eines Tumorregisters im Rahmen auführlicher Modellrechnungen gesondert behandelt (Abschn. 6), deren Ziel es sein wird, konkrete Vorhersagen zu erstellen und diese einer praktizierten Nachsorgestrategie gegenüberzustellen. In diesem Abschnitt wird auf die Frage der Veränderlichkeit der Prävalenz unter den Bedingungen eines Nachsorgeprogramms eingegangen (vgl. Tab.3.8, S.71).

- (1): Grundbedingung zur Teilnahme an einem Nachsorgeprogramm ist das vermutete Vorliegen der Grundkrankheit. Ein sich ändernder Anteil von bezüglich des Grundleidens falsch positiv eingestuften Patienten verändert die Inzidenz von Folgezuständen, da Patienten, die in Wirklichkeit nicht tumorkrank sind, auch nicht an Rezidiven erkranken. Möglicherweise ist der Anteil falsch positiver Tumordiagnosen höher als vermutet. Beispielsweise werden für die Diagnose Mammakarzinom bei Patientinnen unter 30 Jahren Falsch-positiv-Raten von über 20% angegeben (45), wobei die Werte in höheren Altersklassen eher noch ungünstiger liegen sollen. Für andere Tumordiagnosen werden in (45) noch höhere Falsch-positiv-Raten zitiert. Die ständige Verfeinerung der Primärdiagnostik, z.B. durch Markerbestimmung oder immunologische Typisierung, kann die Rate der irrtümlich als Tumorträger diagnostizierten Patienten senken. Dies entspräche einem scheinbaren Rückgang der Quote der Dauerheilungen und zöge ein Ansteigen von Prävalenzen in der Tumornachsorge nach sich.
- (2): Verschiedene Subentitäten einer Tumordiagnose können prognostisch und im Metastasierungsmuster erheblich voneinander abweichen. Trends der prognostischen Faktoren verändern Inzidenz und möglicherweise auch die Wachstumsrate, also die durchschnittliche Dauer von Zielereignissen in der Nachsorge. Dieser Aspekt ist u.a. deshalb von Bedeutung, weil über

Tab. 3.8: Faktoren, die die Prävalenz latenter Tumorprogression in der Nachsorge beeinflussen

Faktor	Wirkung
1. Vorliegen des Grundleidens Krebs	Senkung des Anteils der fälschlich als Tumorpatienten Angesehenen reduziert die Quote der "Dauerheilungen"
2. Staging- und therapieunabhängige prognostische Faktoren	Verschiebungen im prognostischen Spektrum, z.B. durch Krebsvorsorge, bewirken Änderungen der Verlaufseigenschaften
3. Empfindlichere Staging-Diagnostik bei Primärtherapie	Durch verfeinerte Primärdiagnostik wird ein höherer Anteil generalisierter Erkrankungen bereits primär erfaßt und der prophylaktischen Nachsorge entzogen
4. Wirksamkeit der Primärtherapie	Die laufende Verbesserung der Quote von Dauerheilungen ist der Schwerpunkt der klinischen Forschung
5. Definition von Tumorfreiheit	Je strenger die Kriterien an "Tumorfreiheit" nach Primärtherapie definiert werden, um so geringer ist die erwartete Rezidivrate unter den Patienten, die diese Kriterien erfüllen
6. Schichtspezifische Inanspruchnahme	Unterschiedliche Teilnahme, unterschiedliches Spontanverhalten bei Beschwerden. Teilnahme korreliert negativ mit Risiko
7. Ausfall durch natürliches Absterben	Das natürliche Absterben bewirkt "Verjüngung" der länger in Nachsorge Stehenden
8. Verlaufsbedingte Verweigerung	Patienten mit schwierigem Verlauf, die mit dem bisherigen Verlauf nicht "zufrieden" sind, bleiben der weiteren Nachsorge eher fern
9. Frequenz und Intensität der Nachsorge	Je intensiver die Überwachung, desto geringer die Ausbeute pro Untersuchung
10. Veränderung der Ereignisfolge durch intensive Tests	Ein neuer, empfindlicherer Test könnte die bisherigen Vorstellungen zum Krankheitsverlauf in Frage stellen
11. Früherkennung langsam progredienter Ereignisse	Reduktion der Spätrezidive durch einen dem Length-Bias vergleichbaren Effekt

ihn ein direkter Zusammenhang zwischen Krebsfrüherkennungsprogrammen und Tumornachsorgeprogrammen hergestellt wird. Durch Intensivierung von Früherkennungsmaßnahmen steigt möglicherweise der Anteil prognostisch günstiger verlaufender Fälle an (length bias). Zusätzlich kann sich der Anteil der 'unechten' Fälle, deren Metastasierungsrisiko aufgrund der Ausdehnung (Carcinoma in situ) und/oder der Morphologie (Borderline-Tumoren) gering ist, erhöhen. Der Übergang vom günstigen Prognosekriterium zur Fehldiagnose (s.o.) ist fließend.

Fließend ist aber auch der Übergang vom 'prognostischen Faktor' zum 'Nachweis' der erfolgten Generalisierung. Dies wurde bereits in Abschnitt 3.4.2 diskutiert. Ein sensitiveres Erkennen ungünstiger prognostischer Faktoren, wie z.B. in (7), kann analoge Wirkungen auf die Prävalenz haben, wie die im nächsten Punkt zu besprechende Staging-Diagnostik.

- (3): Die Mechanismen, die zu späterer Tumorprogredienz nach therapiefreiem Intervall führen, sind nicht vollständig aufgeklärt. Anzunehmen ist, daß zumindest ein großer Teil der späteren Rezidiv-Befunde auf eine Generalisierung der Tumorerkrankung bereits bei Diagnosestellung zurückzuführen ist. Ziel der Staging-Untersuchungen ist die diagnostische Verifizierung einer bereits vollzogenen Generalisierung mit möglichst hoher Wahrscheinlichkeit, um ggf. auf ultraradikale lokale Eingriffe zu verzichten und/oder geeignete systemische Maßnahmen einzuleiten.

Mit zunehmend empfindlicherer Staging-Diagnostik - als Beispiel seien monoklonale Antikörper genannt - wächst der Anteil der - berechtigt oder ungerechtfertigt - primär als generalisiert eingestuften Krankheitsfälle, die der späteren prophylaktischen Betreuung im Rahmen eines Nachsorgeprogramms teilweise entzogen sind. Hierdurch kann die Prävalenz latenter Tumorprogressionen in der Nachsorge sinken.

- (4): Art und Umfang der Primärtherapie entscheiden häufig über die Prognose des Patienten. Vor allem die adjuvante Therapie zur Beseitigung von klinisch (noch) nicht manifesten Tumor-

resten durch lokale (bezogen auf ein späteres Rezidiv) oder systemische (bezogen auf spätere Fernmetastasierung) Maßnahmen gewinnt zunehmend an Bedeutung. Dies kann, wie oben dargestellt, zu Änderungen der Behandlungsabläufe führen, die sich unmittelbar auf die Metastasenprävalenz in der ambulanten Nachsorge durch niedergelassene Ärzte auswirken. Zusätzlich senken erfolgreiche adjuvante Maßnahmen konzeptbedingt die Inzidenz von späteren Tumorprogressionen.

Andererseits wird befürchtet, daß die Rate der Zweitmalignome für die so behandelten Patienten ansteigen könnte. Auch Zweitmalignome sind Zielereignisse der Tumornachsorge. Ein möglicher Effekt geänderter Therapiekonzepte könnte daher sein, daß Nachsorgeprobleme in Zukunft langfristiger anzugehen sind und evtl. sogar erst später im Verlauf relevant werden.

- (5): Mit zunehmend empfindlicherer Diagnostik stellt sich die Frage nach der Definition von Tumorfreiheit. Dies wurde bereits weiter oben diskutiert. Als extremes Beispiel eines empfindlichen diagnostischen Staging-Tests wurde der axilläre Lymphknotenstatus beim Mammakarzinom genannt. Mehr als 3 positive Lymphknoten deuten auf Generalisierung der Krankheit hin, auch wenn aktuell keine direkten Anzeichen für das Vorliegen von Fernmetastasen bestehen. Für die Verlaufsbeschreibung des Mammakarzinoms insgesamt mag es bedeutungslos erscheinen, ob Patientinnen mit mehr als 3 positiven Lymphknoten primär als tumorfrei eingestuft werden oder nicht. Bezüglich der Metastasenprävalenz in einem Nachsorgeprogramm kann diese Frage jedoch entscheidend sein, sofern das Nachsorgeprogramm an 'tumorfreie' Patienten gerichtet ist.

Angenommen, die genannte prognostisch ungünstige Gruppe würde in Zukunft einheitlich in onkologischen Zentren oder Schwerpunktpraxen adjuvant chemotherapiert und diagnostisch überwacht, so würde die Metastasenprävalenz der für ein allgemeines Nachsorgeprogramm verbleibenden prognostisch günstigeren Zielgruppe reduziert, u.U. auf Bruchteile des Ausgangswertes.

Durch solche Umstellungen wären Nachsorgeprogramme betroffen, die prognoseunabhängig für alle Patienten gleiche Nachsorgestrategien vorschreiben. für diese Programme ergäben sich reduzierte prädiktive Werte. Wurde das Nachsorgeprogramm dagegen an Hand des Lymphknotenstatus stratifiziert, schlagen sich derartige Umschichtungen nicht unmittelbar in den (schichtspezifischen) Ergebnissen nieder.

Die Population, an die sich das Angebot von Tumornachsorgeprogrammen richtet, ist also ständigem Wandel unterworfen. Die Annahme, man könne die Ausgangsbedingungen für Nachsorgeprogramme über längere Zeit konstant halten, ist verfehlt. Welcher Anteil dieser variierenden Zielpopulationen durch das Nachsorgeprogramm jedoch tatsächlich erreicht wird, hängt von der Inanspruchnahme des Versorgungsangebots durch die Betroffenen ab. Auch diese unterliegt einer Reihe von Einflußfaktoren.

- (6): Unabhängig von Nachsorgeprogrammen hängt die Prognose des Individuums von dessen eigenem Gesundheitsbewußtsein ab. Vorsorgemaßnahmen erreichen erfahrungsgemäß gerade die High-Risk-Gruppen am schlechtesten. In (5) wird versucht, die Korrelation von Sozialstatus und Inanspruchnahme von Präventionsleistungen durch die Angebotsstruktur im Gesundheitswesen zu erklären. Entscheidendes Problem sei das "Prinzip der Patientensteuerung", welches das direkte Anbieten von notwendigen Leistungen nicht zulasse und es vom Laien abhängig mache, welche Leistungen er in Anspruch nehme.

Durch die Einführung von Tumornachsorgepässen und Monitoring der Patienten in Nachsorgeregistern wird dieses Prinzip durchbrochen. Dies muß jedoch möglicherweise durch einen anderen Nachteil erkauft werden. Eine strikte Organisation der Nachsorge könnte im Einzelfall Anlaß sein, auftretende Symptome bis zum nächsten Nachsorgetermin, eventuell über Monate, zu übergehen.

Hierbei könnte die Wahrscheinlichkeit, aus eigenem Antrieb frühzeitig zum Arzt zu gehen und bei Beschwerden den nächsten Termin nicht abzuwarten, partiell vom Gesundheitsbe-

wußtsein und damit direkt von der sozialen Schicht des Betroffenen abhängen. So gesehen hätte schichtabhängige Selektion u.U. den Effekt der Prävalenzanreicherung für die regulären Nachsorgeuntersuchungen, bedingt durch Aufschub von erkennbar notwendigen Arztbesuchen zwischen zwei Terminen, d.h. 'Späterkennung' statt Früherkennung von Ereignissen.

(7): Patienten höheren Alters haben eine geringere Wahrscheinlichkeit, an der Nachsorge über den vollen Zeitraum teilnehmen zu können. Betrachtet man alle Tumordiagnosen und beide Geschlechter gemeinsam, so erhöht sich der Altersmedian der tatsächlich in Nachsorge stehenden Patienten innerhalb der ersten 5 Nachsorgejahre statt um 5 Jahre nur unwesentlich um ca. 1 Jahr (vgl. Abschn. 6). Wenn Alter und tumorabhängige Prognose korreliert sind, so sind Rückwirkungen auf die Prävalenz in der Nachsorge nicht auszuschließen.

(8): Die Bereitschaft, den Empfehlungen des Arztes Folge zu leisten, hängt unter anderem von den bisherigen Erfahrungen des Patienten ab. Für Patienten, deren Primärbehandlung und/oder bisheriger Verlauf in der Nachsorge nicht 'programmgemäß' verlief, können erhöhtes Risiko des Auftretens von Rezidiven oder Metastasen und erhöhte Wahrscheinlichkeit, die weitere Zusammenarbeit mit der Medizin zu verweigern, koinzidieren.

Die dritte wesentliche Ursache für Verschiebungen der Prävalenz ist das Nachsorgeprogramm selbst. Einige Einzelaspekte sind zu unterscheiden.

(9): Der Einfluß der Frequenz von Nachsorgeuntersuchungen auf die Prävalenz pro Untersuchung ist trivial. Die Gründlichkeit der Untersuchungen, z.B. Anzahl simultan durchgeführter Tests, wirkt in derselben Richtung.

(10): Bevor ein Ereignis klinisch manifest wird, hat es eine unterschiedlich lange Latenzzeit durchlaufen, in der es durch geeignete Tests mit einer gewissen Wahrscheinlichkeit er-

kennbar gewesen wäre. Nicht unbedingt muß jenes Ereignis, welches ohne Nachsorge als erstes klinisch manifest wird, auch durch Nachsorge als erstes erkennbar werden. Beispielsweise ein hinsichtlich einer bestimmten Lokalisation empfindlicher neu eingeführter Test kann u.U. alle bis dahin gültigen, an anderen diagnostischen Möglichkeiten und damit anderen Zielereignissen orientierten Vorstellungen zum natürlichen Krankheitsverlauf inadäquat erscheinen lassen.

Im Grunde bewirkt die Einführung neuer Test demnach keine Änderung der Prävalenz eines Ereignisses, sondern eine implizite Neudefinition des Ereignisses selbst. Diese Testabhängigkeit der Zielereignisse wurde bereits im Abschnitt 3.4.2 angesprochen und braucht hier nicht erneut diskutiert zu werden. Es sei jedoch darauf hingewiesen, daß wegen der Testabhängigkeit der Zielereignisse eine Extrapolation von aus anderen Datenquellen geschätzten Prävalenzen auf ein bestimmtes Nachsorgeprogramm mit abweichender Diagnostikstrategie nur unter der Proportionalitätsannahme vollzogen werden kann, daß sich durch Nachsorge gemäß der geplanten Strategie zwar die Diagnosezeitpunkte vorverlagern, jedoch nicht die Reihenfolge der Ereignisse verändert wird. In der Diskussion der Modellrechnungen zum Mammakarzinom im 6. Abschnitt wird hierauf zurückzukommen sein.

- (11): Unter Length-Bias wird der Effekt verstanden, daß langsamer wachsende, prognostisch eventuell günstigere Tumoren durch periodische Routineuntersuchungen mit größerer Wahrscheinlichkeit entdeckt werden als schneller progrediente Prozesse. Dieser Bias tritt im Screening und auch in der Nachsorge auf. Jedoch besteht zwischen beiden der Unterschied, daß die Entstehungswahrscheinlichkeit für Primärtumoren - abgesehen von Alterungseffekten - gleichverteilt ist, während der Entstehungszeitpunkt für Rezidive bzw. Fernmetastasierung im Prinzip vor dem Datum der Primärtherapie liegt. Welche Mechanismen anschließend für die Zeitverteilung des Auftretens dieser Ereignisse verantwortlich sind, ist nicht in allen Punkten bekannt. Die Annahme erscheint plausibel, daß sowohl ein langsames, nahezu kontinuierliches Tumor-

wachstum als auch die plötzliche 'Explosion' eines bis dahin 'schlummernden' Tumorrests für Spätereignisse verantwortlich sein können. Spätereignisse aufgrund langsam und kontinuierlich wachsender Rezidivprozesse würden durch Früherkennung in der Nachsorge eliminiert. Damit korrespondiert der Anstieg der Prävalenz in der Frühphase der Nachsorge mit einem Rückgang in den folgenden Jahren.

4. Ansätze zur Bewertung des medizinischen Nutzens von Nachsorgeprogrammen

Die technische Perfektion eines Nachsorgeprogramms, d.h. die Präzision, mit der die eingesetzten Tests im gegebenen Testumfeld arbeiten, ist ein wichtiger Parameter. Wenn die Präzision eines Tests bzw. der prädiktive Wert des positiven Befundes hinter den Erwartungen zurückbleiben, können sich weitergehende Fragen erübrigen. Dennoch garantiert Präzision allein nicht den medizinischen Nutzen von Nachsorgeprogrammen. Es ist zu fordern, daß die frühere Erkennung von Rezidiven, Fernmetastasen und Zweitmalignomen dem Patienten erkennbare Vorteile bietet. Die Frage nach dem medizinischen Nutzen von Nachsorge stellt sich um so dringlicher, je aufwendiger, belastender und invasiver die benutzten Tests sind und je engmaschiger die Nachsorgeuntersuchungen terminiert werden.

In diesem Abschnitt werden unterschiedliche Studienansätze zur Bewertung von Nachsorgeprogrammen dargestellt und bezüglich ihrer Aussagekraft und Realisierbarkeit diskutiert.

4.1 Ex-ante Bewertung ohne empirische Daten

Für zahlreiche Tumorerkrankungen hat sich die Prognose in den letzten Jahrzehnten kaum verbessert. Dies gilt insbesondere dort, wo Stahl und Strahl bisher nicht durch wirksame systemische Therapien ergänzt werden konnten. Auch unter intensiver Nachsorge wird derzeit für diese Tumoren keine erhebliche Verbesserung der Prognose erwartet, so daß sich die empfohlene Nachsorge häufig auf allgemeinärztliche Betreuung ohne eingreifende Diagnostik beschränkt. Im Vordergrund stehen dann die Verhinderung akuter Komplikationen, die Behandlung unspezifischer, teilweise von der Tumorthherapie abhängiger Befindungsstörungen und die klinische Beobachtung des Verlaufs (29). Für diese Form der Nachsorge ist die Frage nach dem medizinischen Nutzen insofern entschärft, als sie den Patienten kaum belastet und geringe Kosten verursacht. Vorrangig stellt sich für solche Maßnahmen die Frage nach der Qualität der Durchführung. Bewertung des Nachsorgeprogramms redu-

ziert sich dementsprechend auf Überwachung der protokollgemäßen Betreuung und Registrierung der erhobenen Befunde. Eine empirische Rechtfertigung dieser Art von ärztlicher Betreuung wird kaum für notwendig erachtet.

Gegenstand der folgenden Überlegungen sind vorwiegend apparativ aufwendigere, risikoreichere und in ihren Zielen anspruchsvollere Nachsorgeprogramme, deren Stellenwert durch a priori angestellte Überlegungen nicht vollständig abgeklärt werden kann.

Beispielsweise reicht wohl auch die internistisch-onkologische Maxime, daß die Chancen einer Chemotherapie mit wachsender Tumormasse zurückgehen, als alleinige Begründung für umfangreichere Nachsorgeprogramme häufig nicht aus. Sie wird durch den für die Behandlung zahlreicher Tumoren noch unklaren Stellenwert der Chemotherapie relativiert. Durch die statistische Auswertung geeigneter Beobachtungen könnte in diesen Fällen zur Klärung offener Fragen beigetragen werden.

4.2 Zielkriterien zur empirischen Evaluation

Obwohl der Arzt in seinen Entscheidungen zahlreiche Randbedingungen zu berücksichtigen hat und meist mehr als ein Ziel im Auge behalten muß, wird beim Vergleich unterschiedlicher Therapien in Studien in der Regel mit einem einzigen oder einer geringen Zahl von Zielkriterien gearbeitet. Die Festlegung nur eines Zielkriteriums ist insbesondere dann vernünftig, wenn entweder die übrigen Randbedingungen und Zielsetzungen unter alternativen, in der Studie zu prüfenden Therapien invariant bleiben oder das gewählte Zielkriterium von überragender Bedeutung ist, z.B. eine mögliche Dauerheilung.

Neben Remissionsdauern und Überlebenszeiten werden aber in jüngerer Zeit verstärkt auch andere, schwerer meßbare Parameter, insbesondere 'Lebensqualität' in unterschiedlichen Ausprägungen in Planung und Analyse klinischer Studien einbezogen. Vor allem wird dies angestrebt, wenn geringfügiger Zugewinn an Überlebenszeit durch aggressive Therapie erkaufte werden muß.

Versuche zur Algorithmisierung des Trade-offs zwischen Qualität und Quantität des Überlebens Tumorkranker werden beispielsweise in (22) beschrieben. Vorgeschlagen wird u.a. die Identifizierung von x Tagen schmerzfreien Überlebens mit $x + y$ Tagen eines Lebens unter Beschwerden. Die Gewichtung von x und y soll hierbei auf der Basis des Urteils des Betreuers oder anhand der Wertvorstellungen des Patienten bzw. potentiellen Patienten (= Gesunden) erfolgen. Solche Modelle wurden mehrfach entwickelt, aber offenbar noch nirgends eingesetzt.

Die Probleme sind offensichtlich. Insbesondere vom Betroffenenurteil abgeleitete 'utility curves' zur Bewertung unterschiedlicher Überlebensdauern unterschiedlicher Qualität (41) sind der Frage ausgesetzt, ob eine Konstanz der Wert- und Zielvorstellungen eines Patienten über den Verlauf seiner Krankheit nicht als Fiktion anzusehen ist.

Was mit formalen Methoden erreichbar ist, ist wohl in erster Linie die weitgehende Objektivierung einzelner Kriterien, um zu gewährleisten, daß individuelle Wertentscheidungen und Trade-offs unter möglichst realistischen Annahmen ablaufen. Unter dieser Prämisse werden im folgenden mögliche Zielkriterien für die Bewertung des medizinischen Nutzens von Nachsorge diskutiert (vgl. Tab. 4.1, S.81).

(1): Bei der Bewertung der Krebsvorsorge spielt die Mortalität eine dominierende Rolle, sowohl unter experimentellen Designs als auch im Rahmen von Interventionsstudien. Ein Rückgang der tumorspezifischen Mortalität wird als (mögliche) Folge einer durch Screening verbesserten Dauerheilungsrate angesehen. Zumindest kann unter der Voraussetzung einer konstanten Inzidenz geschlossen werden, daß die Tumorkrankheit häufiger als bisher bis zum Eintreffen anderer Todesursachen unter Kontrolle gehalten werden konnte.

Durch Nachsorge kann dagegen kein spürbarer Rückgang der tumorspezifischen Mortalität erwartet werden, da die durch Nachsorge ausgelösten Therapiemaßnahmen meist nur palliativ wirksam sind und höchstens zu einer geringen, wenn auch im Einzelfall begrüßenswerten Lebensverlängerung führen. Über

Tab. 4.1: Zielkriterien zur Evaluation des Nutzens systematischer Tumornachsorge

Kriterium	als Maß für	Probleme
1. Mortalität	Verbesserung der Prognose	Kein sensibles Maß bei bloßer Lebensverlängerung; Confounding: Trends, Krebsvorsorge, Therapie
2. Überleben ab Diagnosestellung des Primärtumors	Verbesserung der Prognose	Individuelles Follow-up ab Diagnosestellung erforderlich; Confounding: Trends, Krebsvorsorge, Therapie; bedingt eliminierbar
3. Überleben ab Progression	Relevanz von Nachsorge; (Verbesserung der Prognose)	Length-Bias, Lead-Time-Bias; Confounding: Trends, Therapie; bedingt eliminierbar
4. Karnofsky-Index usw.	"globale" Bewertung der Lebensqualität	Relevanz (Validität); Reliabilität; Confounding: Therapie
5. Spezifische Details, z.B. WS-Spontanfrakturen, Hospitalisierungsrate	Bewertung der Lebensqualität anhand von Details	Datenlage, Indizes
6. Trefferrate	Qualität der Durchführung; Risiko/Nutzen	Validierung; Relevanz
7. Nebenbefunde	Qualität der Durchführung; Risiko/Nutzen	Validierung; Relevanz
8. Materieller Aufwand	Wirtschaftlichkeit	setzt Annahmen über die direkten Wirkungen voraus und ist erst anschließend anwendbar

eine minimale Altersverschiebung hinaus sind daher keine Effekte auf die Mortalitätsstatistik zu erwarten. Der zahlenmäßige Rückgang der Todesursachen selbst ist gering und ergibt sich aus der Wahrscheinlichkeit, in dem kurzen Zeitraum, um den die Nachsorge das Leben verlängert, an einer anderen Ursache zu sterben. Bei einer durch Nachsorge erzielten Lebensverlängerung um 1 Jahr würde der Rückgang beispielsweise nur etwa 1,5 bis 3% betragen, entsprechend der Wahrscheinlichkeit eines 60- bis 70jährigen, innerhalb eines Jahres zu sterben.

Dies setzt zusätzlich voraus, daß Todesfälle aufgrund anderer Ursachen durch die Therapie nicht beschleunigt oder vermehrt auftreten. Derartige Effekte könnten z.B. entstehen, wenn die durch Screening bzw. Nachsorge frühzeitig behandelten Patienten einer aggressiveren Therapie zugeführt werden, die mit höherer Wahrscheinlichkeit als herkömmliche Behandlungsformen zu Zweitmalignomen führt. Die isolierte Beobachtung nur einer spezifischen Todesursache in der Mortalitätsstatistik kann deshalb unzureichend sein.

Die simultane Wirkung mehrerer Faktoren auf die Mortalitätsstatistik (Confounding) ist bei der Interpretation von Mortalitätsdaten schwer zu kontrollieren. In (42) wird beispielsweise über Probleme berichtet, die Hysterektomierten als Confounder in die Analyse der Assoziation von Screening der Cervix und Mortalität an Gebärmutterhalskrebs einzubeziehen. Die Studie fand keinen befriedigenden Abschluß, weil repräsentative Daten offenbar nicht gewonnen werden können.

Die wichtigsten Confounder bei der Bewertung von Tumornachsorge anhand der Mortalitätsstatistik sind Therapie (-änderungen) sowie Trends bezüglich der Inzidenz, bezüglich prognostischer Faktoren oder bezüglich der Teilnahme an der Krebsvorsorge. Die durch Nachsorge in der Mortalitätsstatistik auszulösenden Veränderungen sind in ihrer Größenordnung den durch diese Faktoren bewirkten Effekten unterlegen. Dies erschwert zusätzlich die Evaluation von Nachsorge anhand der Mortalitätsstatistik.

Die Mortalitätsstatistik scheint die Evaluation populationsbezogener Krebsvorsorgeprogramme und anderer präventiver Leistungen ohne individuelles Record-linkage zu gestatten. Im Zusammenhang mit der Reduktion von Confounding verkehrt sich aber gerade dieser vermeintliche Vorzug ins Gegenteil. Deshalb sind Mortalitätsdaten trotz ihrer leichten Verfügbarkeit zur Evaluation von Nachsorgeprogrammen schlecht geeignet.

- (2): Präziser ist die Beobachtung der Überlebenszeit ab Primär-diagnose. Erfüllt ist sowohl die Bedingung, daß das Kriterium bei allen Patienten und unabhängig von der Art der Nachsorge beobachtet wird, als auch die Forderung, daß unter der Nullhypothese der Wirkungslosigkeit von Nachsorge, d.h. der Konstanz des Endpunkts (=Tod), keine Wechselwirkung zwischen Nachsorge und Zielkriterium besteht.

Zwar sind auch bei der Interpretation dieser Maßzahl Confounder zu berücksichtigen, doch kann dies durch das Record-linkage von Einzelfalldaten geschehen. Im Gegensatz zur Mortalität ist dieser Parameter inzidenzunabhängig. Der Einfluß unterschiedlicher prognostischer Faktoren, unterschiedlicher Therapie und unterschiedlichen Screeningverhaltens kann durch Randomisation kontrolliert oder durch Post-hoc-Stratifikation bzw. multivariate Modelle ansatzweise eliminiert werden. Vorausgesetzt wird hierbei die Existenz eines Tumoregisters, in dem prognostische Faktoren prospektiv erfaßt werden und Daten zum Follow-up verfügbar sind.

- (3): Demgegenüber ist die Überlebenszeit ab Diagnose der Progression als Zielkriterium mit Lead-Time-Bias und Length-Bias belastet. Mit Lead-Time wird die Zeit bezeichnet, die aufgrund der Screening- bzw. Nachsorgemaßnahmen durch frühere Entdeckung im Vergleich zur späteren Diagnosestellung anhand von Symptomen gewonnen wurde. Lead-Time wird zum Lead-Time-Bias, wenn im Vergleich von unterschiedlich nachgesorgten Kollektiven Zielkriterien herangezogen werden, die durch diese Lead-Time verändert werden, wie z.B. die Überlebenszeit ab Diagnosestellung der Progression.

Mit Length-Bias wird die Tatsache bezeichnet, daß die Wahrscheinlichkeit, bei einer prophylaktischen Untersuchung entdeckt zu werden, von der Wachstumsgeschwindigkeit des Tumors abhängt: rasch wachsende haben eventuell nie, langsam wachsende möglicherweise mehrmals die Chance, durch Screening bzw. Nachsorge entdeckt zu werden. Damit weisen die im Rahmen von Routineuntersuchungen diagnostizierten Fälle im Durchschnitt ein langsames Wachstum und deshalb möglicherweise eine günstigere Prognose auf als ein nicht überwacht Kollektiv. Der Fehler würde weiter verstärkt, wenn die im Programm nicht frühzeitig entdeckten, sondern anhand von Symptomen manifest gewordenen Fälle in ein durch Nachsorge nicht überwacht Vergleichskollektiv mündeten.

Eine sinnvolle Anwendung wäre die Abschätzung der Relevanz von Nachsorgemaßnahmen. Ungeachtet der Biases könnte die systematische Beobachtung der Überlebenszeit ab Progression für unterschiedliche Untergruppen Hinweise auf die Größenordnung der Effekte geben, die überhaupt durch Nachsorge erreichbar sind.

Zwar existieren zur Prognose des Rezidivs auch außerhalb von Nachsorgeprogrammen zahlreiche Publikationen, ein Defizit an Informationen besteht aber, wenn die Prognose vom Zeitpunkt des Beginns der Metastasierung und eventuell von primären prognostischen Faktoren abhängig gemacht wird. Hierzu würde wiederum der Rückgriff auf die primären Befunde erforderlich. Die Überlebenszeit ab Diagnose der Progression kann deshalb nur ergänzend und nicht anstelle der Überlebenszeit ab Primärdiagnose benutzt werden.

- (4): Für die Evaluation der Wirkung von Tumornachsorge auf die Lebensqualität stehen zahlreiche Skalen zur Verfügung, wobei davon auszugehen ist, daß durch ein Nachsorgeprogramm simultan widersprüchliche Wirkungen auf die Lebensqualität induziert werden können. Hebung und Beeinträchtigung der Lebensqualität durch die Nachsorge sind daher gegeneinander abzuwägen. Dies könnte durch Auflistung einzelner Teilkriterien der Lebensqualität erfolgen, wie z.B. Dauer der Hospitalisierung, Schmerzfreiheit usw..

Versucht man, über diesen Ansatz hinausgehend die unterschiedlichen Aspekte der Lebensqualität in einem einzigen Score zu vereinigen, so stellen sich die bekannten Probleme der Gewichtung einzelner Bereiche. Die große Zahl existierender Scores belegt, daß 'Hebung' der Lebensqualität nicht verbindlich definiert ist.

Bekannt und verbreitet ist der Karnofsky-Index (27) zur Beschreibung des Allgemeinzustandes. Gleichzeitig ist das der älteste einschlägige Index.

PERFORMANCE STATUS		
CONDITION	%	COMMENTS
A. ABLE TO CARRY ON NORMAL ACTIVITY AND TO WORK. NO SPECIAL CARE IS NEEDED.	100	NORMAL. NO COMPLAINTS. NO EVIDENCE OF DISEASE.
	90	ABLE TO CARRY ON NORMAL ACTIVITY. MINOR SIGNS OR SYMPTOMS OF DISEASE.
	80	NORMAL ACTIVITY WITH EFFORT. SOME SIGNS OR SYMPTOMS OF DISEASE.
B. UNABLE TO WORK. ABLE TO LIVE AT HOME, CARE FOR MOST PERSONAL NEEDS. A VARYING DEGREE OF ASSISTANCE IS NEEDED.	70	CARES FOR SELF. UNABLE TO CARRY ON NORMAL ACTIVITY OR TO DO ACTIVE WORK.
	60	REQUIRES OCCASIONAL ASSISTANCE, BUT IS ABLE TO CARE FOR MOST OF HIS NEEDS.
	50	REQUIRES CONSIDERABLE ASSISTANCE AND FREQUENT MEDICAL CARE.
C. UNABLE TO CARE FOR SELF. REQUIRES EQUIVALENT OF INSTITUTIONAL OR HOSPITAL CARE. DISEASE MAY BE PROGRESSING RAPIDLY.	40	DISABLED. REQUIRES SPECIAL CARE AND ASSISTANCE.
	30	SEVERELY DISABLED. HOSPITALIZATION IS INDICATED ALTHOUGH DEATH NOT IMMINENT.
	20	HOSPITALIZATION NECESSARY, VERY SICK. ACTIVE SUPPORTIVE TREATMENT NECESSARY.
	10	MORIBUND. FATAL PROCESSES PROGRESSING RAPIDLY.
	0	DEAD.

Abb. 4.1: Karnofsky-Index zur Beschreibung des Allgemeinzustandes (27)

Entsprechend der überwiegend somatischen Orientierung wird der Karnofsky-Index u.a. als Entscheidungskriterium zur Wahl der Therapie bzw. als Steuerungs- und Abbruchkriterium eingesetzt. Andererseits wird die Einstellung des Patienten zum Leben nur lückenhaft reflektiert. Beispielsweise besteht keine Korrelation zwischen dem Karnofsky-Index und 'Glück', während 'Treppensteigen' und 'Balancehalten' gut mit dem Index korrelieren (72).

Wegen dieser vorwiegend somatischen Ausrichtung des Karnofsky-Index wurden zahlreiche alternative Indizes entwickelt. Als Beispiel sei der QL-Index nach (59) wiedergegeben, der die Aspekte Aktivität, alltägliche Verrichtungen, Gesundheit, Unterstützung durch die Umwelt und Einstellung zum Leben umfaßt.

QUALITY OF LIFE INDEX

Study No. _____

Age _____ Sex M / F (ring appropriate letter) □□

Primary Problem or Diagnosis _____ □

Secondary Problem or Diagnosis or complication (if appropriate) _____ □□□

Scorer's Specialty _____ □□□

SCORING FORM

Score each heading 2, 1 or 0 according to your most recent assessment of the patient

ACTIVITY	<i>During the last week the patient</i> * has been working or studying full time or nearly so in usual occupation or managing own household or participating in unpaid or voluntary activities, whether retired or not. 2 * has been working or studying in usual occupation or managing own household or participating in unpaid or voluntary activities, but requiring major assistance or a significant reduction in hours worked or a sheltered situation or was on sick leave. 1 * has not been working or studying in any capacity and not managing own household. 0	□
DAILY LIVING	<i>During the last week the patient</i> * has been self reliant in eating, washing, toileting and dressing, using public transport or driving own car. 2 * has been requiring assistance (another person or special equipment) for daily activities and transport but performing light tasks. 1 * has not been managing personal care nor light tasks and/or not leaving own home or institution at all. 0	□
HEALTH	<i>During the last week the patient</i> * has been appearing to feel well or reporting feeling great most of the time. 2 * has been lacking energy or not feeling entirely up to par more than just occasionally. 1 * has been feeling very ill or lousy, seeming weak and washed out most of the time or was unconscious. 0	□
SUPPORT	<i>During the last week</i> * the patient has been having good relationships with others and receiving strong support from at least one family member and/or friend. 2 * support received or perceived has been limited from family and friends and/or by the patient's condition. 1 * support from family and friends occurred infrequently or only when absolutely necessary or patient was unconscious. 0	□
OUTLOOK	<i>During the past week the patient</i> * has usually been appearing calm and positive in outlook, accepting and in control of personal circumstances. 2 * has sometimes been troubled because not fully in control of personal circumstances or has been having periods of obvious anxiety or depression. 1 * has been seriously confused or very frightened or consistently anxious and depressed or unconscious. 0	□
QL INDEX TOTAL		□□

How confident are you that your scoring of the preceding dimensions is accurate? Please ring the appropriate category

Absolutely Confident 1	Very Confident 2	Quite Confident 3	Not Very Confident 4	Very Doubtful 5	Not at all Confident 6
---------------------------	---------------------	----------------------	-------------------------	--------------------	---------------------------

□

Abb. 4.2: QL-Index zur Messung der Lebensqualität chronisch Kranker nach (59)

Eine Übersicht über Verfahren und Probleme der Reliabilitäts- und Validitätsbestimmung von zur Messung der Lebensqualität geeigneten Scores wird in (22) gegeben.

- (6): Von einer guten ärztlichen Betreuung von Tumorkranken in der Nachsorge dürfte erwartet werden, daß sie sich auf solche Indizes der Lebensqualität auswirkt. Zusätzlich ist, vor allem für differenziertere Nachsorgestrategien mit organspezifischer apparativer Diagnostik, die Beobachtung spezieller Einzelkriterien von Vorteil, deren Auswahl sich aus den Verlaufseigenschaften der Zielerkrankung ergibt. Beispielsweise könnte für das Mammakarzinom und andere Tumorarten mit hoher Metastasierungswahrscheinlichkeit in das Skelett die Rate der Spontanfrakturen ein aussagekräftiges Kriterium zum Vergleich unterschiedlicher Ansätze in der Nachsorge darstellen.

Auch die immateriellen Kosten eines Nachsorgeprogramms können als Synopse von Einzelkriterien dargestellt werden, wie Belastung durch die Teilnahme, testbedingte Komplikationen sowie Begleitumstände und Folgen der Abklärung falsch-positiver Befunde.

- (7): Das Kriterium der Trefferraten sei nur der Vollständigkeit halber aufgeführt. Durch Trefferraten werden technische Aspekte beschrieben, die nicht unmittelbar mit medizinischem Nutzen gleichzusetzen sind. Allenfalls sind niedrige Trefferraten als inverser Nutzen anzusehen. Auf die Gefahr der Vermengung der Effekte von Nachsorge und regulärer, ohnehin erfolgreicher Betreuung bei der Ermittlung von Trefferraten sei hingewiesen. Vergleichbare Erfahrungen existieren aus der Evaluation der Krebsfrüherkennungsprogramme, deren Befunde bekanntlich nur zum Teil an asymptomatischen Patienten erhoben werden.

Auch die Rate der in Nachsorgeprogrammen erhobenen Nebenbefunde läßt nur bedingt auf den Nutzen der Maßnahmen schließen. Aus Screeningprogrammen ist bekannt, daß die Rate von Nebenbefunden unterschiedliche Interpretationen zuläßt. Sinnlose und kostspielige Steigerung der Nachfrage nach Behandlung ist eine, kostengünstigere und wirksamere Betreuung zur sekundären Prävention von Erkrankungen eine andere Interpretation.

Die Objektivität der Beobachtung könnte durch den Versuch der Rechtfertigung zusätzlicher Diagnostik- bzw. Therapie-maßnahmen überlagert sein. Auch die Erfassung der 'objektiven' Tumorparameter im Verlauf der Nachsorge ist grundsätzlich von diesem 'Rechtfertigungs-bias' betroffen. Im Unterschied zu den Nebenbefunden hilft bei der Rezidiv- und Metastasensuche jedoch die Beobachtung des späteren Krankheitsverlaufs. Ein 'Rückfall in die reguläre Nachsorge nach einem als verdächtig oder pathologisch dokumentierten Befund ermöglicht eine angemessene Interpretation des Vorgangs und damit die Elimination des Bias. Was zurückbleibt, ist eine erkennbare Verschiebung der Spezifität eines Tests aufgrund einer Überkapazität an der diagnostischen Einrichtung, wie

Für die Einschätzung des Stellenwerts der erhobenen Nebenbefunde ist ein solches Regulativ nicht vorhanden. Insofern handelt es sich bei der Rate von Nebenbefunden um ein Erfolgskriterium mit unbekanntem Vorzeichen. Dennoch würde durch einen vollständigen Verzicht auf die Erfassung und Evaluation von Nebenbefunden u.U. ein markanter Teil der Versorgungsrealität vernachlässigt. Bei der Festlegung möglicher Zielkriterien für eine Nachsorgestrategie wird darauf zu achten sein, daß vorwiegend relevante, reliabel erfassbare und anhand geeigneter Unterlagen nachträglich in Stichproben überprüfbare Nebenbefunde berücksichtigt werden.

- (8): Die Festlegung der Kosten, die einem Nachsorgeprogramm direkt zurechenbar sind, ist möglich. Der Informationswert dieser Berechnung ist jedoch gering, solange nicht zumindest die quantifizierbaren und ökonomisch adäquat bewertbaren Effekte des Programms gegenübergestellt werden, z.B. Ersparnisse an Pflegeaufwand, Mehraufwendungen für oder Ersparnisse an Chemotherapie oder positive und negative Auswirkungen von Nebenbefunden.

So gesehen entspricht die ökonomische Analyse des Programms der Transposition medizinischer Kriterien in ein anderes Bezugssystem. Dementsprechend kann das verfügbare Instrumentarium der Kosten-Nutzen-Analyse erst zur Anwendung kommen, wenn die direkten Nutzenkriterien beschreibbar sind.

Die hier diskutierten Kriterien messen oder beschreiben unterschiedliche Aspekte des individuellen Nutzens in unterschiedlichen Dimensionen. Ihre Verfügbarkeit stellt, wie auch die Schätzung von Sensitivität, Spezifität und prädiktivem Wert, Anforderungen an die Organisation von Nachsorgeprogrammen und die zugehörige Dokumentation, wobei durch den Langzeitcharakter besondere Probleme aufgeworfen sind.

Einige dieser Kriterien erfordern zudem bestimmte Analysetechniken, so z.B. die Veränderung der Mortalitätsstatistik, während andere in ihrer Verwendbarkeit weniger eingeschränkt sind. Das Kriterium der Lebensqualität wäre beispielsweise als Zielgröße einer Kohorten-/Interventionsstudie ebenso wie zur Definition von

Fällen und Kontrollen im Rahmen einer Fall-Kontroll-Studie benutzbar. Fall-Kontroll-Studien erfordern hierbei Dichotomisierung. So ist beispielsweise der Übergang von der Überlebenszeit in Tagen zur x-Jahres-Überlebenszeit zwar mit Informationsverlust verbunden, die Charakteristika der Kriterien, wie sie oben diskutiert wurden, bleiben jedoch durch solche Transformationen unberührt.

Aufgrund der Komplexität des Problems sind zur Bewertung von Nachsorgeprogrammen Details zur Primärtherapie, sowie Fakten über die Inanspruchnahme der Leistungen und zum Nachsorge- und Krankheitsverlauf erforderlich. Die wichtigsten der oben diskutierten Zielkriterien sind langfristig, sie weisen über das Ende des tumorfreien Intervalls hinaus. Im Kontrast dazu muß der Hinweis eingeführter Nachsorgeprogramme gesehen werden, daß die Nachsorge mit dem Eintritt einer Progression endet und Daten nur über die Durchführung der Nachsorge selbst im Zeitraum des tumorfreien Intervalls gesammelt werden. Damit werden die wichtigsten Bewertungskriterien bereits im Ansatz ausgeblendet.

4.3 Randomisierte Studien im Vorfeld der Breitenversorgung

4.3.1 Festlegung der zu prüfenden Alternativen

Zuerst ist davon auszugehen, daß mit der Evaluation differenzierter Nachsorgestrategien multiple Fragestellungen verbunden sind. Wann (bzw. wie oft), womit, wo, für wen, evtl. auch durch wen sind simultane, gleichwertige Fragen an ein Programm. In einem kontrollierten Ansatz könnten demgegenüber nur wenige Schemata verglichen werden. Aus der möglichen Vielfalt müßten daher einige Strategien hinsichtlich der genannten Kriterien fixiert werden. Festzulegen wären u.a. der zeitliche Abstand von Untersuchungen, die Wahl von Zielgebieten und die Auswahl oder Kombination von Tests mit gleichem Zielgebiet. Für differenzierte Strategien kann zusätzlich die Festlegung von Untergruppen erforderlich werden.

Ähnliche Probleme stellen sich bei der Definition von Therapiestudien in der Onkologie, wo aus der Vielzahl wirksamer Substanzen und publizierter Schemata einige Kombinationen ausgewählt und

nach Festlegung der Details zur Dosierung, Applikation, Behandlungsdauer etc. verglichen werden. Der Vergleich der Zielkriterien gibt Anhaltspunkte zur Bewertung der Alternativen.

Zur Bewertung alternativer Nachsorgestrategien könnte ähnlich vorgegangen werden. Allerdings wäre aufgrund des Langzeitcharakters der Intervention und des u.U. sehr unterschiedlichen Aufwands für verschiedene Leistungen innerhalb einer Nachsorgestrategie das Bedürfnis nach weitergehenden Aussagen höher. Anliegen der Sekundäranalyse wäre die Auflösung des Paketcharakters der angewandten Strategie. Auf welchen Faktor bezogen sich beobachtete Effekte? War ein bestimmter Test, ein bestimmtes Zielgebiet, eine bestimmte Untergruppe hauptsächlich für beobachtete Effekte verantwortlich, war es die Wechselwirkung zwischen zwei Faktoren? Aus klinischer Sicht würden erst solche Detailaussagen als wirklich informativ empfunden, vergleichbar dem Übergang zum multiplen Testen nach Ablehnung einer globalen Nullhypothese beim statistischen Testen. Gerade diese Sekundäranalyse hätte jedoch überwiegend deskriptiven bzw. explorativen Charakter, auch wenn die Beobachtungen in einer randomisierten und korrekt durchgeführten Studie gesammelt wurden.

Analoge Probleme der Sekundäranalyse wurden bereits für die Analyse von Screeningprogrammen diskutiert. Wie problematisch selbst in der überschaubaren Situation der Krebsvorsorge die Interpretation in Untergruppen ist, wird aus der Analyse und Diskussion der HIP-Studie ersichtlich (48). Dies wurde im Abschn. 2.3. referiert.

Die Alternative, Nachsorgeprogramme miteinander zu vergleichen, die sich nur genau in einem einzigen Parameter voneinander unterscheiden, darf als utopisch betrachtet werden. Der Einblick, der auf diesem Wege in den komplexen Mechanismus Nachsorge gewonnen werden könnte, wäre angesichts des für eine randomisierte Studie erforderlichen hohen Aufwandes zu gering. Jede der oben genannten Strategiekomponenten, wie Frequenz, Organe, Tests oder Untergruppen kann in gleicher Weise einer Frage würdig sein. Die Fiktionierung aller Komponenten bis auf eine wäre ethisch und erkenntnistheoretisch schwer zu vertreten. Allein aus diesem Grunde könnte die oben beschriebene explorative Sekundäranalyse nicht umgangen werden.

4.3.2 Durchführbarkeit der Studie

Daneben stellt sich die Frage, ob angesichts der Vielfalt der Parameter und der Länge des Interventionszeitraums (bis zu 5 Jahre) eine definierte Strategie überhaupt an hinreichend vielen Personen hinreichend konstant applizierbar ist. Welche Abweichungen sind noch tolerabel, dürfen die drop-out-Fälle überhaupt als Zufälle angesehen werden usw. - dies sind Fragen, die sich zusätzlich aus der pragmatischen Bewertung des Ablaufs der Studie stellen müßten. Schon bei kurzzeitigen Therapiestudien wirft die protokollgerechte Durchführung gelegentlich Probleme auf, für eine mehrjährige Intervention durch Nachsorge sind diese vermehrt zu erwarten. Vor der statistischen Auswertung müßten post-hoc-Klassifikationen erfolgen, z.B. zur Definition der "korrekten" Teilnahme, durch die die Auswertung, unabhängig von den oben genannten Faktoren, von der konfirmatorischen auf die explorative Seite geschoben würde.

Neben den Schwierigkeiten der Auswertung und Interpretation solcher Studien ergeben sich auch Probleme der Durchführung der Randomisierung und des Monitoring. Randomisation als Zuteilung von Nachsorgestrategien auf Individuen wäre vorwiegend in größeren Zentren durchführbar. Unter den Bedingungen der ambulanten kassenärztlichen Versorgung wäre die korrekte Durchführung der Randomisation problematisch. Allein die unterschiedliche ökonomische Bedeutung alternativer Leistungen in den zu prüfenden Studienarmen wäre, noch vor allen logistischen Problemen, ein Hindernis.

Beschränkung auf größere Zentren hätte andererseits den unerwünschten Effekt, daß die Ergebnisse, die dort bezüglich Durchführbarkeit und Präzision des Programms ermittelt wurden, nicht unmittelbar auf die Breitenversorgung übertragbar wären. Hinzu käme aber, daß auch in größeren Zentren die Anwendung unterschiedlicher Langzeitstrategien nach Randomisierungsplan aus organisatorischen Gründen schwer durchführbar wäre. Gleichzeitig zögen Studien in großen Zentren die Notwendigkeit nach sich, nicht nur Problemfälle, sondern alle dort behandelten Tumorpatienten zur Nachsorge an die Klinik zu binden. Dies wäre nur realisierbar, wenn den niedergelassenen Ärzten auf Jahre die Durchführung von Tumornachsorge vorenthalten oder erschwert würde.

4.3.3 Veränderlichkeit der Randbedingungen

Die Übertragbarkeit von Studienergebnissen auf zukünftige Verhältnisse ist, unabhängig von Studiendesign und Qualität der Durchführung, an die Konstanz der Randbedingungen gebunden, unter denen die Studie durchgeführt wurde.

In der Tumornachsorge wirken diesbezüglich drei Faktoren limitierend zusammen, und zwar

- die lange Dauer der Intervention Nachsorge,
- die teils geringe Konstanz der Primär- und Rezidivtherapie sowie
- die Rate technischer Innovationen in der Diagnostik.

Die Auswirkungen therapeutischer und diagnostischer Innovationen auf die Nachsorge werden im folgenden anhand zweier Beispiele diskutiert.

Therapeutische Neuerungen wurden während der vergangenen Jahre in besonderer Weise beim nichtseminomatösen malignen Hodentumor erzielt. Bis zur Einführung der zytostatisch wirksamen Substanz CIS-Platin galt die Prognose dieses Tumors als ungünstig. Heilungsaussichten bestanden nur in Ausnahmefällen. Der Stellenwert von Nachsorge war in dieser Phase wegen des Fehlens wirksamer Rezidivtherapie gering. Der Einsatz von CIS-Platin führte dagegen zur erheblichen Verbesserung der Überlebenschancen von Tumorpatienten, auch bei generalisierter Erkrankung. Traditionelle Therapiekonzepte wurden ersetzt.

Im Tumorzentrum München wurde in dieser Phase eine kontrollierte Studie über den Stellenwert der Chemotherapie zur adjuvanten Behandlung des malignen Teratoms im Stadium IIa geplant. Verglichen werden sollte die präoperative Kurzzeitbehandlung mit einer adjuvanten Chemotherapie. Der Verzicht auf adjuvante Maßnahmen war im Studienprotokoll nicht vorgesehen.

Noch in der Planungsphase dieser Studie wurde jedoch die Wirksamkeit der neuen CIS-Platin-haltigen Chemotherapie mehrfach reproduziert, so daß nach Abschluß der Planung auf die Durchführung der Studie verzichtet wurde. Stattdessen wurde im Vertrauen auf

die bei Progression anwendbare Chemotherapie auf adjuvante Maßnahmen gänzlich verzichtet.

Zur Absicherung dieses Vorgehens werden seither die betroffenen Patienten im Rahmen der Nachsorge besonders häufig und intensiv diagnostisch überwacht, so daß nötigenfalls frühzeitig interveniert werden kann. Nachsorge ist zu einer aktiven Komponente im Therapiekonzept aufgerückt. Eine Extrapolation dieser Entwicklung auf die nächsten Jahre erscheint gewagt. Vorstellbar wäre jedoch, daß mit wachsendem Vertrauen in die Wirksamkeit der internistischen Behandlung des Hodentumors nach der adjuvanten Therapie auch die intensive Nachsorge allmählich wieder reduziert werden könnte.

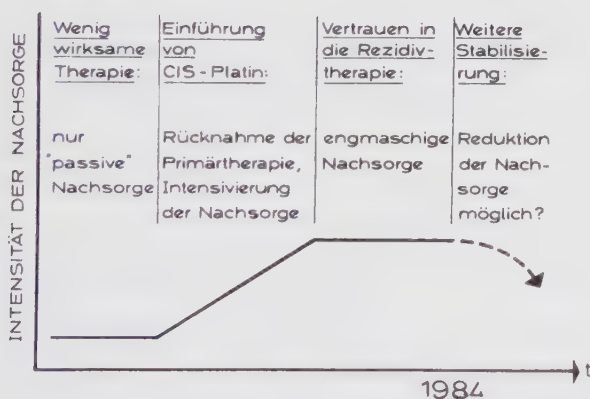


Abb. 4.3: Wandel des Stellenwerts von Tumornachsorge aufgrund von Therapieinnovationen am Beispiel des malignen Hodentumors

Dieses Beispiel zeigt, daß Tumornachsorge nach zwei Seiten hin im therapeutischen Gesamtkonzept verankert ist. Sowohl Art und Umfang der Primärtherapie als auch die Handlungsmöglichkeiten beim späteren Auftreten einer Progression sind ausschlaggebend für den Stellenwert von Nachsorgemaßnahmen.

Auch Verbesserungen und Neuerungen in der Diagnostik schaffen neue Voraussetzungen für die Nachsorge. Jüngstes Beispiel innovativer Technik in der Diagnostik bildet die Kernspintomographie (NMR). Zwar sind die Angaben über den Indikationsbereich und die

Aussagekraft der neuen Methode noch widersprüchlich. In Anknüpfung an die Entwicklung der Computertomographie oder die kontinuierliche Ausweitung der Anwendungsgebiete für Sonographie kann aber vermutet werden, daß sich auch durch die NMR Einbrüche in die Anwendungsbereiche anderer diagnostischer Tests ergeben werden (18). Dies müßte zu Änderungen der meisten bestehenden Nachsorgeempfehlungen führen. Auch nach Einführung und Verbreitung der NMR dürfte dieser Prozeß nicht beendet sein.

Diese geringe Konstanz der Diagnostik und Therapie behindert die empirische Bewertung von Nachsorge. Um ein Nachsorgeprogramm mit einer Laufzeit von 5 Jahren nach Primärtherapie abschließend beurteilen zu können, müssen hinreichend viele Patienten dieses Programm zuzüglich einer angemessenen Beobachtungszeit durchlaufen haben. Dieser Zeitraum erhöht sich um die Rekrutierungszeit, die Planungsphase der Studie und zuletzt die Auswertung. Frühestens 8 Jahre nach Definition einer Nachsorgestrategie kann daher mit Ergebnissen prospektiver Studien gerechnet werden. Für die Umsetzung als Breitenmaßnahme wären anschließend u.U. zusätzliche Zeiten anzusetzen, so daß erst etwa 8 bis 10 Jahre nach Definition einer Nachsorgestrategie mit deren Anwendung in einem breitenwirksamen Nachsorgeprogramm gerechnet werden könnte, wollte man den prospektiv erfolgten Nachweis des Nutzens abwarten.

Der so ermittelte Zeitbedarf steht im Kontrast zur 'onkologischen Halbwertszeit' differenzierter und eng mit der Therapie verzahnter Nachsorgestrategien. Lediglich Basisprogramme für prognostisch ungünstige und therapeutisch schwer beeinflussbare Tumorerkrankungen könnten u.U. in diesem zeitlichen Rahmen geprüft werden, ohne an Stabilität zu verlieren. Für Basisprogramme stellt sich jedoch diese Notwendigkeit am wenigsten, wie eingangs festgestellt.

Eine ex ante erfolgende Überprüfung des medizinischen Nutzens von Nachsorgeprogrammen trifft damit auf Probleme der fehlenden Konstanz der Voraussetzungen und auf Grenzen der Erkennbarkeit des richtigen Verhaltens durch pragmatische Limitierungen.

4.4 Methodische Ansätze zur Evaluation in der Breitenversorgung

Aufgrund dieser Limitierungen der Aussagekraft und Durchführbarkeit randomisierter Studien ist die Möglichkeit in Frage gestellt, vor der Breitereinführung eines Nachsorgeprogramms vollständig über dessen Wirkungen informiert zu sein. Eine Etablierung umfangreicher Nachsorgeprogramme unter Mißachtung dieses Informationsdefizits würde als unbefriedigend empfunden. Jedoch kann auch die extreme Alternative, die Unterlassung jeder Art von Nachsorge, nicht befriedigen, da zumindest 'Basisaktivitäten', insbesondere die Möglichkeit zur regelmäßigen klinischen Untersuchung und allgemeinärztlichen Betreuung, von unmittelbarem Vorteil für jeden Tumorpatienten sind. Damit ist die Aufgabe gestellt, einen Weg zwischen den Extremen zu finden, der angesichts unvollständiger Evidenz verfügbare Informationen nutzt und zur allmählichen Erweiterung des Wissens führt. Erfordernisse und Perspektiven werden nachfolgend diskutiert.

4.4.1 Interpretation säkularer Trends

Die Analyse säkularer Trends zur Evaluation einer Intervention ist angezeigt, wenn Daten über Zielkriterien und/oder Intervention nur in aggregierter Form verfügbar sind und die Zusammenführung auf Fallebene nicht erreicht werden kann. Typisches Beispiel ist die Evaluation eines Krebsvorsorgeprogramms durch Korrelation von Inanspruchnahme und Mortalitätsstatistik. Der tatsächliche Krankheitsverlauf für die im Früherkennungsprogramm entdeckten Fälle ist nicht bekannt, auch für die am Tumor erkrankten Nicht-Teilnehmer liegen keine Verlaufsdaten vor. Änderungen der Sterbewahrscheinlichkeit werden als Indiz für das Vorliegen der gewünschten Wirkung der Intervention Krebsvorsorge angesehen.

Die mit dieser Art der Evaluation verbundenen Limitierungen sind bekannt. Neben dem allgemeinen Problem der Führung von Kausalitätsnachweisen durch Interventionsstudien stößt die Interpretation globaler, nicht auf Individuen beziehbarer Trends zusätzlich auf die Schwierigkeit, daß die Koinzidenz von Intervention und Wirkung am Individuum nicht einmal überprüfbar ist.

Mögliche Zielkriterien zum medizinischen Nutzen von Tumornachsorge wurden in Tabelle 4.1 genannt. Zur Analyse säkularer Trends bietet sich aus dieser Liste vorwiegend die Mortalitätsstatistik an, da keines der anderen Kriterien global registriert wird. Allerdings sind, wie oben erwähnt, gerade die Auswirkungen auf die Mortalitätsstatistik durch Tumornachsorge vergleichsweise gering.

Unter der Voraussetzung einer geeigneten Organisationsstruktur und ausreichender Informationsflüsse bietet der Mikrokosmos Tumornachsorge im Gegensatz zu den laufenden Krebsvorsorgeprogrammen die Chance, Daten über Intervention, Zielereignis und weitere Faktoren ad Individuum zusammenzuführen und mit diesen Daten zu einer akzeptablen Evaluation zu kommen. Von ausschlaggebender Bedeutung für solche Ansätze ist die Kooperation mit einem Tumorreger, da ein Teil der Zielkriterien, z.B. Überleben ab Primär-diagnose, sowie die meisten Schichtkriterien, u.a. Art und Umfang der Primärtherapie und prognostisch relevante Faktoren, nur so verfügbar werden. Durch diesen Ansatz besteht die Möglichkeit, für die im Register erfaßten Patienten beobachtete Assoziationen zwischen Nachsorge und Zielkriterium mit Hilfe zusätzlicher Daten zu interpretieren.

Andererseits werden weder jetzt noch in Zukunft die prognostisch relevanten Faktoren, nicht einmal Inzidenzen, flächendeckend in Tumorregerstern erfaßt. Im Gegenteil, je differenzierter das Spektrum der im Register verarbeiteten Informationen ist, um so enger wird der Ausschnitt der erfaßten Population. Damit stellt sich evtl. das Problem, die aus der Verbindung von Tumorregerster und Nachsorgeverlauf gewonnene Evidenz auf die nicht im Einzugsbereich des Registers befindliche Bevölkerung zu extrapolieren. Für diese Extrapolation könnte die Beobachtung globaler Maßzahlen und deren Entwicklung im zeitlichen Verlauf einen gewissen Stellenwert erreichen. Hierbei wäre insbesondere an aus der laufenden Nachsorge abgeleitete Indikatoren wie Teilnahmeraten, Befundhäufigkeiten usw. zu denken.

4.4.2 Fall-Kontroll-Studien

Fall-Kontroll-Studien werden durchgeführt, um auffällige Assoziationen zwischen dem Zielereignis und möglicherweise kausalen Faktoren zu identifizieren. Ausgangspunkt ist das eingetretene Zielereignis. Von Zielereignis Betroffene, die sogenannten Fälle, werden mit geeigneten Kontrollen, also Personen, an denen das Zielereignis nicht beobachtet wurde, verglichen. Testgröße des Vergleichs ist die Odds-Ratio, d.h. der Quotient aus den Odds (Anzahl der Exponierten/Anzahl der Nichtexponierten) der Fälle und der Kontrollen. Die Odds-Ratio wird als Maß für das mit der Exposition verbundene Risiko, das Zielereignis zu entwickeln, benutzt.

Auf die Technik der Fall-Kontroll-Studien kann unter anderem zurückgegriffen werden, wenn eine experimentelle Beeinflussung des diskutierten Auslösers der beobachteten Wirkung nicht in Betracht kommt, wie dies im Fall der Tumornachsorge häufig anzunehmen sein wird.

Im folgenden wird vorausgesetzt, daß ein definiertes Nachsorgeprogramm als allgemein verfügbare Leistung angeboten wird, daß die Teilnahme an einer Nachsorgeuntersuchung zentral erfaßt wird und daß Daten zum gleichen Patienten zusammengeführt sowie gegebenenfalls mit zugehörigen Daten in einem Tumorregister verknüpft werden können. In Abschnitt 5.3 wird ein Konzept zur personenbezogenen Verlaufsbeobachtung der Nachsorge bei Wahrung des Datenschutzprinzips kurz skizziert. In dieser Struktur kann Nachsorge, gegebenenfalls in Verbindung mit einem Tumorregister, als prospektive Kohortenstudie betrachtet werden. Fall-Kontroll-Studien innerhalb von Kohortenstudien werden als Hybrid-Design bezeichnet. Chancen und Probleme solcher hybrider Designs zur Evaluation des medizinischen Nutzens eines Nachsorgeprogramms werden im folgenden diskutiert.

Im Rahmen laufender Nachsorgeprogramme bieten sich zahlreiche Möglichkeiten zur Unterscheidung von Fällen und Kontrollen anhand der Krankheitsverläufe an. Fälle könnten z.B. durch den Lifestatus x Jahre nach Primärtherapie, durch das Vorhandensein einer bestimmten Komplikation etc. definiert werden, wobei als Kontrol-

le der jeweils komplementäre Teil der in Frage kommenden Patienten zur Verfügung steht. Als Exposition wäre die mehr oder weniger regelmäßige Teilnahme an der angebotenen Nachsorge anzusetzen.

Die Eigenschaft des "Infragekommens" bedarf hierbei der Präzisierung. Voraussetzung zur Aufnahme einer Fall-Kontroll-Studie ist

- die Unabhängigkeit von Exposition und Rekrutierungswahrscheinlichkeit sowie
- die Verfügbarkeit der zur Stratifizierung benötigten Basisdaten sowie des Zielkriteriums.

Um die Unabhängigkeit von Exposition und Rekrutierung zu gewährleisten, muß die Erfassung von Fällen und Kontrollen von der laufenden Nachsorge unabhängig sein. Dies erfordert die prospektive Erfassung von Tumorpatienten zum Zeitpunkt der Primärtherapie und damit ein Tumorregister, aus dem Fälle und Kontrollen zu rekrutieren sind.

Mit der Existenz eines Tumorregisters ist auch hinsichtlich der Verfügbarkeit von Basisdaten eine notwendige Voraussetzung gegeben. Ob diese auch hinreichend ist, kann nur im konkreten Fall einer planenden Studie in einer konkreten Registerumwelt beantwortet werden. Bedeutsam sind insbesondere Arbeitsweise und verarbeitetes Merkmalsspektrum des betreffenden Registers.

Auch unter der Voraussetzung eines arbeitenden Tumorregisters ist die Durchführbarkeit einer Fall-Kontroll-Studie nicht in jedem Fall gewährleistet. Weitere Fragen betreffen die Definition und Entscheidbarkeit des Zielereignisses, die Klassifikation der durchgeführten Nachsorge und die Stratifizierung. Diese und weitere für die Planung und Durchführung hybrider Fall-Kontroll-Studien in der Tumornachsorge bedeutsame Gesichtspunkte werden in der Tabelle 4.2 zusammengefaßt und anschließend diskutiert, soweit sie nicht auf allgemeine methodische Probleme von Fall-Kontroll-Studien zu reduzieren sind.

Tab. 4.2: Fragen und Probleme bei der Durchführung von Fall-Kontroll-Studien zur Bewertung des medizinischen Nutzens eines Nachsorgeprogramms.

Gegenstand	Probleme	mögliche Bias
1. Rekrutierung der Fälle und Kontrollen	Unabhängigkeit von der laufenden Nachsorge	Selection
2. Zielereignis (Definition der Fälle)	Entscheidbarkeit der Fall-Eigenschaft, Definition des Auftretenszeitraums Klassifikation/Schwellenwert Konkurrierende Risiken	Misclassification
3. Klassifikation der Nachsorge ("Exposition")	Gesamtzahl der Untersuchungen Regelmäßigkeit tolerable Abweichungen Untersuchungsmethoden, Vollständigkeit der Untersuchungen pro Termin, zusätzliche Untersuchungen (Frequenz, Organe) Ort der Untersuchung Vermengung mit regulärer Betreuung Ende der Nachsorge bei Progression	Misclassification
4. Schichtung	Prognostische Faktoren Therapie (Klassifikation/Standard?) Zeit bis zum Zielereignis Begleitkrankheiten Sozioökonomisch, Wohnort usw.	Confounding, Interpretation

- (1): Die Notwendigkeit der Unabhängigkeit von Rekrutierung und Nachsorge wurde bereits oben als Grundforderung genannt. Mit einem arbeitenden Tumorregister und klinikweise vollständiger Erfassung von Tumorpatienten wird dieser Forderung entsprochen, auch wenn nicht alle Kliniken mit dem Register kooperieren. Wird die Fall-Kontroll-Studie auf der Basis z.B. des Tumorregisters durchgeführt, so hängt die Wahrscheinlichkeit eines Patienten, als Fall bzw. Kontrolle in die Studie zu gelangen, zwar von der behandelnden Klinik, nicht aber von der Teilnahme an der Nachsorge ab.

- (2): Dagegen ist die Entscheidbarkeit der Fall-Eigenschaft auch unter der Voraussetzung eines arbeitenden Tumorregisters nicht unmittelbar gewährleistet. Je nach Definition des Zielereignisses ergeben sich unterschiedliche Möglichkeiten und Probleme. Befriedigende Ergebnisse sind zu erwarten, wenn das Zielkriterium sich auf den Life-Status des Patienten x Jahre nach Diagnosestellung bezieht. Diese Information kann über ein arbeitendes Tumorregister für nahezu alle Tumorpatienten erhalten werden.

Für die Teilnehmer an einem Nachsorgeprogramm wären darüber hinaus auch alle anderen diskutablen Zielkriterien verfügbar, z.B. Vorliegen von Fisteln, andere somatische Probleme oder auch psychosoziale Aspekte, da jedes dieser Kriterien im Zuge der laufenden Betreuung erhoben werden könnte. Allerdings wären Fall-Kontroll-Studien dann durch mangelnde Beobachtungsgleichheit zwischen Teilnehmern an der Nachsorge und anderen Patienten belastet, denn für Nicht-Teilnehmer wären zur Erfragung des Zielkriteriums Hausbesuche oder briefliche Befragung erforderlich. Nicht zuletzt stieße ein solches Vorgehen auch auf analoge Bedenken wie bei der Diskussion um Haustürinterviews im Rahmen der epidemiologischen Tumorätiologieforschung. Fall-Kontroll-Studien, die auch Nicht-Teilnehmer an der Nachsorge einbeziehen und mit anderen Zielkriterien als Überlebensraten arbeiten, sind daher zurückhaltend zu beurteilen.

Entschließt man sich dennoch zur Verwendung anderer Zielkriterien als des Überlebens, so stellt sich zusätzlich das Problem der konkurrierenden Risiken. Es ist festzulegen, wie z.B. verstorbene Patienten zu berücksichtigen sind, wenn als Zielereignis eine Dichotomisierung des Allgemeinzustandes, das Vorliegen einer Fistel oder ein vergleichbares Kriterium gewählt wird.

In Abschnitt 4.2 wurde gegen die Vermengung konkurrierender Zielkriterien votiert, um die Transparenz von Ergebnissen aufrecht zu erhalten. Übertragen auf die Situation einer Fall-Kontroll-Studie bedeutet dies Vernachlässigung

konkurrierender Risiken, insbesondere Einschränkung der Studie auf die zum fraglichen Zeitpunkt am Leben befindlichen Patienten. Die auf diese Art über Nachsorge zu erzielenden Aussagen sind konditional, die beobachteten Wirkungen der Nachsorge gelten unter der (eventuell ebenfalls nachsorgeabhängigen) Voraussetzung, bis zum fraglichen Zeitpunkt überlebt zu haben.

Die Definition des Zielereignisses wird abgerundet durch die Festlegung von Zeiträumen, Schwellenwerten etc. Erforderlich sind Klassifikationen, die zum Verlust von Information führen. Sinnwidrige Festlegungen können zur Neutralisierung vorhandener Effekte führen, z.B. durch Festlegung einer zu langfristigen Überlebensrate, die trotz erheblicher Prognoseunterschiede weder von Fällen noch von Kontrollen erreicht wird. Vor Abschätzung der mutmaßlich relevanten Bereiche, z.B. Festlegung der Zeit x bei der x-Jahres-Überlebensrate, kann auf die verfügbaren Beobachtungen zurückgegriffen werden. Beispielsweise trifft eine Festlegung des Basiszeitpunkts für die Überlebensrate in der Nähe der medianen Überlebenszeit aller Patienten mit großer Wahrscheinlichkeit den relevanten Bereich sowohl für die Teilnehmer als auch die Nichtteilnehmer am Nachsorgeprogramm.

(3/4): Die Vielschichtigkeit der Expositionsvariablen "Nachsorgeverhalten" wurde bereits im Zusammenhang mit der Diskussion experimenteller Ansätze erörtert (vgl. 4.3.2). Die absehbare Notwendigkeit zur Durchführung einer Post-Hoc-Klassifikation der Teilnahme an der Nachsorge wurde als eine der wesentlichen Limitierungen für experimentelle Ansätze erkannt.

Fall-Kontroll-Studien sind in diesem Punkt flexibler. Explorative Ansätze sind möglich, sowohl zur Auffindung geeigneter Klassifikationen der Expositionsvariablen als auch bezüglich der Stratifizierung. Der Alpha-Fehler kann hierbei durch Mehr-Sample-Techniken kontrolliert werden. Eine Übersicht über diese für epidemiologische Studien typischen Techniken liegt z.B. in (28) vor.

Hybride Fall-Kontroll-Studien zur Bewertung des medizinischen Nutzens von Nachsorgeprogrammen sind also möglich, wenn geeignete Zielkriterien und Auswertungsstrategien gewählt wurden. Nicht zuletzt wird die Durchführbarkeit solcher Studien von den verfügbaren Fallzahlen abhängen, so daß die Größe des Einzugsgebiets des benutzten Tumorregisters von entscheidender Bedeutung ist. Je größer die durch dieses Register aktivierte Fallzahl ist, um so breiter wird der Spielraum für explorative Ansätze.

4.5 Modellbildung zur Planung und Bewertung von Nachsorgeprogrammen

Die Abgrenzung des "mathematischen Modells" gegenüber anderen Evaluationstechniken ist unscharf und willkürlich. Auch statistische Tests sind als mathematische Modelle anzusehen. Im folgenden werden als Modelle aufwendige mathematische Verfahren zur Beschreibung oder Optimierung von in der Regel multifaktoriellen Systemen bezeichnet. Beispiele sind in der folgenden Tabelle zusammengestellt.

Tab. 4.3: Beispiele mathematischer Modelle, die zur Planung oder Evaluation von Nachsorgeprogrammen geeignet sein könnten.

Ansatz	Modell	Arbeitsweise/mögliche Anwendung
Deskriptiv/ analytisch	COX (12)	Regressionsmodell für inkomplette Zeitdauerdaten; Integration von prognostischen Faktoren, Therapie, Nachsorge und Prognose
Deskriptiv	TALLIS (63)	Schätzung von Wachstumsraten; Schätzung des Wachstums von Rezidivtumoren anhand von Diagnosezeitpunkt des Rezidivs, Art der Befundssicherung und Zeitpunkt des letzten tumorfreien Befundes
Konstruktiv	EDDY (20)	Optimierung periodischer Untersuchungen; komplexes Modell eines gesamten Nachsorgeprogramms
Konstruktiv	DANTZIG (13)	Lineare Optimierung; kostenbezogene Ansätze

Konstruktive Modelle sind durch ihre Abhängigkeit von Vorgaben bzw. Parametern gekennzeichnet. Ist die empirische Basis schmal, wie dies bei praktischen Anwendungen meist der Fall ist, so sind die Ergebnisse von Modellrechnungen mit großer Unsicherheit belastet. Welche Limitierungen sich hieraus ergeben, wird in Abschnitt 6.3.4.2 diskutiert.

Probleme werden auch von den Urhebern konstruktiver Modelle selbst gesehen. In (21) wird festgestellt: "Die Probleme der Datenlage sollten nicht unterschätzt werden. Die Theorie ist einfach und die zur Schätzung der Parameter benötigten Daten können ohne Schwierigkeit beobachtet werden, jedoch sind derzeit diese Daten im allgemeinen nicht verfügbar".

Nachsorgeprogramme, durch die Verlaufsinformationen nach definierten Beobachtungskriterien (Zeitraster, diagnostische Tests, etc.) verfügbar werden, können entscheidend zum Abbau des Datendefizits beitragen. Im Zusammenhang mit Fragen der Erweiterung des Leistungsangebotes, beispielsweise aufbauend auf ein auf klinische Untersuchung und allgemeinärztliche Betreuung beschränktes Basisnachsorgeprogramm, könnten solche Modelle erfolgversprechend angewandt werden. Damit wären die aufgrund geplanter Programmodifikationen zu erwartenden Effekte simulierbar.

Deskriptive Modelle finden ihr potentielles Anwendungsgebiet demgegenüber in der Bewertung der bereits erbrachten Leistungen. Die in diesem Zusammenhang auftretenden Probleme entsprechen weitgehend den Schwierigkeiten, die bei der Diskussion der Durchführung von Fall-Kontroll-Studien dargestellt wurden. Daher sind auch für deren Behebung analoge Überlegungen anzustellen.

Jedes mathematische Modell ist mit spezifischen Limitierungen behaftet, weil es die Realität unter simplifizierenden Annahmen beschreibt. Gemeinsames Problem aller diskutierten Evaluationstechniken ist die Verfügbarkeit der benötigten Daten und die Tatsache, daß zur Bewertung von Nachsorgeprogrammen Daten benötigt werden, die nach beiden Richtungen auf der Zeitachse über die Nachsorgeperiode selbst hinausweisen. Die Integration von Daten über Primärbefund, Nachsorge und weiteren Krankheitsverlauf ist daher vorrangig zu lösen.

5. Anforderungen an die Planung und Bewertung von Nachsorgeprogrammen

5.1 Zur Anwendbarkeit der WHO-Kriterien an Screeningprogramme in der Nachsorge

Über Planung und Evaluation von Früherkennungsprogrammen, besonders im Bereich der Krebsvorsorge, liegen weltweit umfangreiche Erfahrungen vor. Die Monographie (71) der WHO nennt 10 Kriterien, im folgenden als "WHO-Kriterien" bezeichnet, die als Voraussetzung zur Etablierung von Screeningprogrammen erfüllt sein sollen bzw. anhand derer bestehende Programme zu überprüfen sind. Zur Beurteilung von Screeningprojekten bzw. -programmen werden diese Kriterien weitgehend anerkannt. Die Anwendbarkeit dieser Punkte auf die Tumornachsorge wird im folgenden auf der Basis der Ergebnisse der Abschnitte 3 und 4 untersucht.

Ein offensichtlicher Unterschied zwischen Screening und Tumornachsorge ist die Multidimensionalität der Zielsetzungen der Nachsorge. Selbst die Bewertung eines in der Zielsetzung eindeutigen Screeningprojektes ist nicht unbedingt in sich widerspruchsfrei, auch wenn methodisch nach dem Stand des Wissens vorgegangen wurde, wie z.B. in der HIP-Studie. Bei der Bewertung eines Nachsorgeprogramms mit seinen vielschichtigen und verästelten Aufgaben muß dementsprechend verstärkt mit divergierenden und widersprüchlichen Einzelaspekten gerechnet werden, die sich allenfalls unter Mühen in ein geordnetes Ganzes einfügen lassen. Die nachfolgende, auf die Probleme der Nachsorge zugeschnittene Modifikation der WHO-Kriterien (Tab. 5.1) versucht, einen Rahmen hierfür vorzugeben.

Tab. 5.1: WHO-Kriterien an Screeningprogramme (71) und Modifikationen für die Tumornachsorge

Screening	Tumornachsorge
1. Relevanz des Gesundheitsproblems für die Gemeinschaft	Frage erübrigt sich im speziellen Fall
2. Existenz anerkannter Behandlungsformen	Konsistenz von Diagnostik und Therapiemöglichkeiten, Transparenz der Wirkungen
3. Strukturelle Voraussetzungen des Versorgungssystems zur Befriedigung des Diagnostik- und Therapiebedarfs	Strukturelle Voraussetzungen des Versorgungssystems zur Befriedigung des Diagnostikbedarfs
4. Existenz eines erkennbaren Frühstadiums	Existenz eines erkennbaren und handlungsrelevanten Frühstadiums
5. Existenz geeigneter Tests	Existenz geeigneter Tests
6. Akzeptabilität des/der Tests	Akzeptabilität und Tragbarkeit der Risiken für die verwendeten Tests
7. Kenntnis des biologischen Verhaltens	Kenntnis des natürlichen Krankheitsverlaufs nach Primärtherapie
8. Übereinstimmung bzgl. Behandlungsbedürftigkeit	Kann nicht gefordert werden
9. Ausgewogenheit von Kosten pro entdecktem Fall und Gesamtaufwand für die medizinische Versorgung	Ausgewogenheit von Kosten pro Tumorpatient und Gesamtaufwand für die medizinische Versorgung
10. Zeitliche Konstanz	Stabile Basisorganisation, Flexibilität gegenüber Strategie-Modifikationen

(1): Sowohl aufgrund der hohen Prävalenz als auch bzgl. der Folgen für den einzelnen Patienten muß die Krebskrankheit als eines der großen Gesundheitsprobleme unserer Zeit angesehen werden. Die insgesamt für die Forschung aufgebrauchten Geldmittel belegen dies. In der Mehrzahl der letal verlaufenden Krankheitsfälle ist erst das spätere Rezidivieren der Krankheit unmittelbar für den schicksalhaften Ausgang.

Sowohl die Tumorkrankheit insgesamt als auch der Krankheitsabschnitt nach der Primärtherapie für sich allein stellen deswegen für die Gemeinschaft relevante Gesundheitsprobleme

dar. Damit ist die Basisfrage nach der Relevanz des Gesundheitsproblems für die Gemeinschaft für den Tumorbereich beantwortet und erübrigt sich daher als Kriterium an Tumornachsorgeprogramme.

- (2): Die Existenz anerkannter Behandlungsformen wird in den WHO-Kriterien als die wesentlichste Grundlage für sinnvolle Screeningmaßnahmen eingestuft. Vor allem wird gefordert, daß die Behandlung der Erkrankung in einem früheren Stadium zu einer nachweisbaren Besserung der Prognose führen soll. Wenn dies nicht von vorherein angenommen werden kann, wird die Klärung im Experiment empfohlen. Solche Experimente wurden für die Früherkennung von Krebskrankheiten stellenweise durchgeführt.

Ergänzend steht hinter allen Krebsfrüherkennungsprogrammen eine gemeinsame Modellvorstellung der zuerst lokalen Tumorentstehung. Der frühzeitigen Entdeckung wird deshalb eine höhere Wahrscheinlichkeit zugeschrieben, die Krankheit in einem lokalen Stadium zu eliminieren und so den Patienten heilen zu können. Mit dieser theoretischen Vorstellung wird die Mehrzahl der existierenden Krebsfrüherkennungsprogramme begründet, soweit experimentelle Daten fehlen.

Zur Bewertung von Tumornachsorge ist der Stellenwert experimenteller Ansätze geringer als bei Screeningprojekten. Dies wurde in Abschnitt 4.3 diskutiert. Auch die Modellvorstellungen zur Begründung von Nachsorge sind weniger zwingend und variieren, wie auch die Ziele von Nachsorge, zwischen den einzelnen Tumoren. Die möglichen Ziele für Nachsorgeprogramme wurden bereits ausführlich diskutiert.

Die Forderung "anerkannter" Therapie entspräche diesen Modellvorstellungen und den Ergebnissen aus Abschnitt 4 nur bedingt. Eine weniger rigide Formulierung ist die Forderung nach Existenz von Therapien, für die begründete oder begründbare Hoffnung auf therapeutischen Nutzen besteht. Diagnostische Überwachung ohne Handlungsmöglichkeit ist unethisch. "Primum nil nocere" bedeutet, Schaden - auch durch Diagnostik - fernzuhalten, wenn Nutzen nicht erwartet werden kann.

An diesen Zielen hat sich die Bewertung der Maßnahmen zu orientieren, zu fordern ist die Konsistenz von Diagnostik und Therapiemöglichkeiten. Die begleitende Beobachtung der relevanten Zielkriterien ist integraler Bestandteil dieser Forderung.

Hieraus leitet sich insbesondere die zusätzliche Forderung der Transparenz der Wirkungen ab. Nachsorgeprogramme sollten transparent sein in dem Sinne, daß Daten zur Bewertung der Zielkriterien laufend verfügbar werden, daß säkulare Trends in der Struktur des nachgesorgten Kollektives meßbar sind und die Selektion der Teilnahme beurteilbar ist. Der Rückgriff auf Patienten soll über die behandelnde Institution möglich sein, Detailfragen sollen durch Spezialerhebungen, z.B. Fall-Kontroll-Studien beantwortbar sein. Für den allmählichen Ausbau von Nachsorgeprogrammen ist die Transparenz des bisher Erreichten als Kernstück zu betrachten.

- (3): Nachsorgeprogramme belasten, ebenso wie Screeningprojekte, den Gesundheitsetat der Gemeinschaft. Allerdings werden nach der gültigen Vorstellung vom spontan selten reversiblen, in der Regel kontinuierlich progredienten Verlauf von Tumorkrankheiten wenig zusätzliche Behandlungsfälle durch Nachsorgeprogramme 'produziert'. Allenfalls kann die frühere Entdeckung einer Progression dazu führen, daß eine abweichende, unter Umständen intensivere Therapie erforderlich bzw. zweckmäßig wird. Sollten sich hieraus Auswirkungen ergeben, so könnten diese auch als Forderung an den Ausbau des Versorgungssystems anstatt als Restriktion an Nachsorgeprogramme formuliert werden.

Dies unterscheidet Nachsorgeprogramme vom Screening an Gesunden, die erst aufgrund des Screeningresultats zu Patienten werden. Beispielhaft ist das Screening auf Cervixkarzinom, das häufig zur Entdeckung potentieller Frühstadien führt, deren biologisches Verhalten nicht prognostizierbar ist. Plausible Schätzungen (1) veranschlagen die screeningbedingten Therapieindikationen auf etwa das zehnfache der Inzidenz invasiver Karzinome, mit entsprechender Erhöhung der Patientenzahlen.

Der steigende Bedarf an diagnostischen Ressourcen könnte dagegen auch für Nachsorgeprogramme entscheidungsrelevant werden, denn dieser ist im Gegensatz zur Häufigkeit von Rezidivbehandlungen durch die Wahl intensiverer oder weniger intensiver Nachsorgestrategien direkt steuerbar. Auch ein belegbarer individueller Nutzen durch Nachsorge rechtfertigt nicht jeden Aufwand, weder ethisch noch ökonomisch. Zu fordern ist deswegen, daß die strukturellen Voraussetzungen des Versorgungssystems zumindest zur Befriedigung des Diagnostikbedarfs ausreichen.

Die Anforderungen, welche ein laufendes Nachsorgeprogramm an die Ressourcen des Versorgungssystems stellt, ergeben sich zum einen aus der Tumorzinzidenz, zum anderen aber aus der gewählten Strategie des Nachsorgeprogramms. Durch Übergang vom Sechsmonatsrhythmus zu dreimonatlichen Untersuchungen wird beispielsweise der zu erbringende Aufwand verdoppelt. Auch die Hinzunahme weiterer Tests hat erhebliche Auswirkungen auf den zu erbringenden Aufwand. Die Durchführbarkeit eines Programms wird deshalb großenteils von der gewählten Strategie bestimmt. Als Kriterium für gegebene oder fehlende Durchführbarkeit könnte daher die Verallgemeinerungsfähigkeit des in einem Programm geforderten Diagnostikaufwands angesehen werden. Hierunter wird verstanden, daß prognostisch und diagnostisch vergleichbare Nachsorgeprobleme für alle Tumorerkrankungen mit gleichem Diagnostikaufwand bedacht werden können, ohne daß dies zur Überforderung der Gesamtkapazität führt. Engpässe durch Verallgemeinerung könnten sowohl bei der Belegung diagnostischer Einheiten als auch bezüglich des insgesamt resultierenden Kostenvolumens auftreten.

Ein Frühstart einzelner tumorspezifischer Nachsorgeprogramme mit nicht verallgemeinerbar hohem, nicht für alle Tumordiagnosen in gleicher Weise realisierbarem Diagnostikaufwand wirft erhebliche ethische Probleme auf, falls sich für solche Programme ein belegbarer Nutzen ergibt.

- (4): Die Existenz eines erkennbaren asymptomatischen Frühstadiums ist auch für Nachsorgeprogramme eine notwendige Forderung.

Allerdings bezieht sich im Fall der "Früherkennung" von Organmetastasen die Bezeichnung Frühstadium nur auf den lokalen Prozeß. Die Krankheit insgesamt ist natürlich bei erfolgter Metastasierung generalisiert, d.h. im fortgeschrittenen Stadium.

Die Bedeutung des Begriffs Frühstadium kann maßnahmenspezifisch variieren. Die Frühzeitigkeit kann sich auf chirurgische oder radiologische Prävention lokaler Komplikationen oder auf die Tumormasse (systemische Behandlung) beziehen. Hypothesen über mehrstufige, kaskadenartige Metastasierungsmechanismen geben sogar dem Postulat einer "frühzeitigen Generalisierung" eine gewisse, zumindest theoretische Berechtigung. Durch Intervention an der ersten Kaskade der Metastasierung könnte die weitere Ausbreitung unterbunden oder verzögert werden. Neben der Erkennbarkeit des Frühstadiums ist explizit die - im Screening durch die Testauswahl a priori garantierte - Verifizierbarkeit von Befunden zu fordern. Tests, deren Ergebnisse - trotz guter Spezifität - nicht zur Rechtfertigung einer Maßnahme ausreichen und die nicht durch Referenzmethoden überprüfbar sind, sind von eingeschränktem Wert. Die Benutzung eines Tests als bloßes Frühwarnsystem bedarf besonders sorgfältiger Begründung.

- (5): Die Existenz geeigneter Tests ist sowohl für Screening als auch für Nachsorgeprogramme zu fordern. Typische Anforderungen an Screeningtests sind Einfachheit der Anwendung und Schnelligkeit der Ausführung. Die Anforderungen hinsichtlich Spezifität treten im Screening in den Hintergrund. Es werden eine oder höchstens zwei Untersuchungen simultan eingesetzt. Dies gilt auch für multiphasisches Screening, denn die dabei zum Einsatz kommenden Tests betreffen disjunkte Zielereignisse und sind voneinander unabhängig. Positive Screeningbefunde werden durch Referenzmethoden diagnostisch abgeklärt.

In der Nachsorge umfaßt diese Forderung die Anwendung sinnvoller Testkombinationen und praktikabler Strategien zur Abklärung auffälliger Befunde. Einfachheit und Schnelligkeit von Tests sind wünschenswert, jedoch nicht Bedingung. Häufig kommen gerade die aufwendigsten und teuersten Tests zur Anwendung.

Typisch für die Nachsorge ist dagegen die Forderung, daß die durch den Test verfügbar werdenden Informationen sinnvoll in therapeutische Maßnahmen umgesetzt werden können. Hierzu müssen überhaupt Therapien verfügbar sein, d.h. Diagnostik- und Therapiemöglichkeiten müssen im Einklang stehen. Außerdem müssen die Aussagen der Tests verifizierbar bzw. hinreichend sicher sein. Diese Gesichtspunkte wurden bereits weiter oben angesprochen. Zusätzlich aber muß der erhaltene Befund lokal bzw. regional hinreichend präzise sein, um den Einsatz der in Frage kommenden Therapie zu ermöglichen. Beispielsweise könnte der Anstieg eines Tumormarkers - Spezifität vorausgesetzt - hinreichend zur Einleitung einer Chemotherapie sein, für lokale oder regionäre Maßnahmen dagegen wäre die Information von geringerem Wert, da der Ort der Tumorprogression nicht ersichtlich ist. In Abschnitt 3.4 wurde dies diskutiert.

- (6): Im genannten Fall wären Markertests als Frühwarnsystem nur Baustein einer Teststrategie. Hierbei stellt sich die Frage der Akzeptabilität für den Patienten, der mit dem Wissen über den positiven Befund belastet wird, ohne daß andere Handlungsmöglichkeiten bestehen, als abzuwarten. Solches Vorgehen bedarf sorgfältiger Begründung. Andererseits rechtfertigt das mit Krebskrankheiten verbundene hohe Risiko für die Betroffenen auch belastende Untersuchungen, sofern deren Nutzen belegbar ist. Für den Betroffenen ist der Nutzen einer Untersuchung nicht abschätzbar. In der Regel vertraut er als Patient der Empfehlung seines Arztes. Nimmt er das Angebot der Nachsorge nicht wahr, so sind die Ursachen in erster Linie nicht in der fehlenden subjektiven Akzeptabilität der verwendeten Tests zu suchen.

Wesentlicher in der Nachsorge ist die Tragbarkeit der Risiken der verwendeten Tests. Die Ziele, die mit einzelnen Nachsorgemaßnahmen verfolgt werden, sind von unterschiedlicher Wertigkeit. Auch die verfügbaren Tests sind unterschiedlich invasiv oder belastend. Beides muß im Einzelfall im Einklang stehen, eingreifende Diagnostik wird nur durch entsprechende therapeutische Perspektiven gerechtfertigt.

Zusätzlich wird das vertretbare diagnostische Risiko durch Prävalenz und/oder Sensitivität bestimmt. Je seltener das Zielereignis bzw. je niedriger die Entdeckungswahrscheinlichkeit, um so größer muß bei gleicher Belastung der therapeutische Gewinn für die wenigen positiven Befunde sein.

Schließlich gehen Häufigkeit und Folgen von falsch-positiven Befunden in die Abwägung zwischen Risiko und erhofftem Nutzen ein. Je größer die durch Unterlassung der Untersuchung ggf. vermeidbare Belastung aufgrund der Abklärung falsch-positiver Befunde, um so größer muß der erhoffte Nutzen eines richtigen positiven Befundes sein. Abzuwägen sind also Risiko der Durchführung, Risiko der Abklärung und Chancen bei richtig-positivem Befund unter Berücksichtigung der jeweiligen Wahrscheinlichkeiten.

- (7): Die Kenntnis des biologischen Verhaltens wirkt für Nachsorgeprogramme weniger Fragen auf als in der Krebsvorsorge, denn für manifeste Tumorerkrankungen ist die Hypothese der stetigen Progredienz kaum umstritten. Ausnahmen werden nur selten mitgeteilt (17). Für die Krebsvorsorge dagegen scheint diese Annahme nicht ohne weiteres gerechtfertigt zu sein, wie durch Modellrechnungen (1) zum Carcinoma in situ der Cervix belegt wird.

Kritisch in der Tumornachsorge ist dagegen die Kenntnis des natürlichen Krankheitsverlaufs. Valide Angaben über die Prävalenz bzw. Inzidenz von Zielereignissen im Verlauf der Nachsorge sind Voraussetzung engmaschiger Nachsorgeprogramme. Je ungenauer die verfügbaren Daten Aufschluß geben, desto geringer ist die Rechtfertigung 'blinder' Intensivdiagnostik. Andererseits können validere Daten nur durch systematisches Handeln gewonnen werden.

- (8): Die Existenz allgemein anerkannter Behandlungsformen wurde nicht als Grundlage von Nachsorgeprogrammen gefordert. Dementsprechend kann auch Übereinstimmung bzgl. der Behandlungsbedürftigkeit nicht in allen Fällen gefordert werden.

Die Therapieplanung onkologischer Erkrankungen wird zugleich durch die Regeln der Heilkunst, soweit solche formuliert sind, und die Würdigung der individuellen Umstände getragen. Die Standardisierung von Therapie und Nachsorge von Krebskrankheiten scheint in beträchtlichem Umfang zur Kollision mit den Gegebenheiten des Einzelfalles zu führen. So verweigern nach den Daten des Tumorregisters München in einzelnen Fächern über 10% der Patienten die ihnen vorgeschlagene Primärtherapie. Gerade für ältere und sehr alte Patienten sind onkologische Standards und individuelle Bedürfnisse teils nur schwer zur Deckung zu bringen. Für Therapieentscheidungen spielen daher subjektive Überlegungen des Arztes eine erhebliche Rolle. Daher ist bei der Beurteilung der Behandlungsbedürftigkeit mit beachtlichen Schwankungen im interinstitutionellen, interindividuellen und vielleicht sogar intraindividuellen Vergleich zu rechnen.

- (9): Zur Ermittlung der Kosten eines Nachsorgeprogramms kann nicht, wie bei Screeningprogrammen, vom entdeckten Fall ausgegangen werden. Dies stünde im Widerspruch zu den multiplen Zielen der Nachsorge, die zu 'Entdeckungen' unterschiedlicher Tragweite führen kann. Für einen trefferorientierten Ansatz müßten unterschiedliche Entdeckungen gewichtet werden.

Da es sich andererseits bei den durch Nachsorgeprogramme insgesamt angesprochenen Personen um einen definierten Kreis von Tumorpatienten handelt, liegt der Bezug auf diese Zahl nahe. Gefordert wird die Ausgewogenheit der Proportion von Kosten des Programms pro Patient aus der Zielgruppe und Gesamtaufwand für die medizinische Versorgung. Natürlich können einem Nachsorgeprogramm ohne Kenntnis von Details keine Kosten zugeordnet werden, da diese je nach Umfang der Maßnahmen variieren.

- (10) Zum Aufbau von Nachsorgeprogrammen sind komplexe Organisationsstrukturen zu schaffen, die nicht ad hoc funktionsfähig werden. Die systematische Nachsorge aller Tumorkranken ist nur mit Hilfe der niedergelassenen Ärzte möglich. Ergebnisse

aus befristeten Follow-up-Studien großer Zentren sind nur eingeschränkt auf die Anwendung in der Breitenversorgung übertragbar. Die kontinuierliche Gewinnung von Langzeitergebnissen zur Bewertung und Verbesserung von Nachsorgestrategien wird in diesem Umfeld nur durch eine langfristig stabile Basisorganisation vorstellbar.

Flexibilität gegenüber Strategiemodifikationen im Detail wird erforderlich, falls Nachsorgeprogramme während der Laufzeit an den veränderten Stand des Wissens angepaßt werden sollen. Zeitliche Konstanz ist schwerer erreichbar als z.B. für ein Screeningprogramm auf Mammakarzinom, da die Tumornachsorge stärker durch therapeutische und diagnostische Innovationen beeinflusst wird.

5.2 Entwurf einer erweiterten Anforderungsliste

Der Übergang von der Planung zur Bewertung eines Nachsorgeprogramms ist, wie bereits eingangs festgestellt wurde, fließend. Daher mag die Zuordnung einzelner Funktionen zu Planung oder Bewertung stellenweise arbiträr erscheinen. Identische Probleme treten in jeweils modifizierter Form mehrfach im Verlauf der Planung und Bewertung eines Programms auf. So bezieht sich die Frage nach der Risiko-Nutzen-Abwägung eines diagnostischen Tests, in der Planungsphase gestellt, auf die verfügbare Evidenz, etwa prognostizierten Nutzen, geschätzte Komplikationsraten und erwartete Inanspruchnahme. Dieselbe Frage im Bewertungskontext zu beantworten, erfordert auswertbare Daten und deren Interpretation. Im Zuge des schrittweisen Ausbaus eines Nachsorgeprogramms gehen beide Aspekte ineinander über.

Die folgende Anforderungsliste verzichtet daher auf die Trennung in Planung und Bewertung, um identisch formulierte Fragen zu vermeiden. Unabhängig davon wurden in den vorausgehenden Abschnitten auf identisch formulierte Fragen unterschiedliche Antworten gegeben, entsprechend dem jeweiligen Kontext.

Tab. 5.2 Entwurf einer erweiterten Anforderungsliste an
Nachsorgeprogramme

1. Vorteile für die Betroffenen
 - 1.1 Bringt Früherkennung durch Nachsorge erkennbare bzw. mutmaßliche Vorteile im Vergleich zum passiven Follow-up?
 - 1.2 In welchen Fällen werden Vorteile gesehen?
2. Vorliegen erkennbarer Frühstadien und Verfügbarkeit diagnostischer Tests
 - 2.1 Ist das gesuchte Frühstadium einer Progression erkennbar und verifizierbar?
 - 2.2 Existenz geeigneter Tests
 - 2.3 Sind die verfügbaren diagnostischen Tests prinzipiell akzeptabel?
 - 2.4 Welche Unannehmlichkeiten und Risiken sind mit der Durchführung der Tests, der Abklärung falsch-positiver Befunde und der definitiven Befundsicherung verbunden?
3. Verhältnismäßigkeit
 - 3.1 Ist die durch einen diagnostischen Test verfügbare Information relevant bezüglich der theoretischen Vorstellungen zum Krankheitsverlauf?
 - 3.2 Ist die durch einen diagnostischen Test verfügbare Information handlungsrelevant?
 - 3.3 Welchen Stellenwert hat der durch den Befund ermöglichte Eingriff?
 - 3.4 Steht der applizierte Diagnostikaufwand mit dem Stellenwert des dadurch ermöglichten Eingriffs im Einklang?
 - 3.5 Wie wahrscheinlich ist es aufgrund der Zusammensetzung der Zielpopulation sowie der Risiko-Nutzen-Abwägung bzw. Akzeptabilität des Eingriffs, daß dieser auch durchgeführt wird?
4. Datenlage bei Planung der Nachsorge
 - 4.1 Wurde die verfügbare Evidenz über Wirkungen und Wirksamkeit der geplanten Maßnahmen aufbereitet und berücksichtigt?
 - 4.2 Wurden Pilotstudien durchgeführt?
 - 4.3 Sind Daten zur Anwendung mathematischer Modelle zur Unterstützung der Planung verfügbar?

5. Prävalenz

- 5.1 Wie wahrscheinlich ist das Vorliegen der diversen Zielereignisse pro Zeitpunkt?
- 5.2 Wurde berücksichtigt, daß sich die Abfolge von Krankheitsergebnissen im Verlauf durch aktive Nachsorge ändern kann?
- 5.3 Wurde der häufig inverse Zusammenhang zwischen Risiko und Inanspruchnahme in Rechnung gestellt?
- 5.4 Wurden weitere Faktoren berücksichtigt, durch die sich die Prävalenz an Fernmetastasen und lokoregionären Rezidiven im säkularen Verlauf ändern kann, insbesondere Wandlung des Krankengutes durch
 - spezifischere bzw. sensitivere Handhabung der Primärdiagnose,
 - verfeinerte Stagingdiagnostik mit wachsendem Anteil primär als generalisiert erkannter Krankheitsstadien,
 - Verschiebungen in den Häufigkeitsverteilungen prognostisch relevanter Faktoren,
 - Änderung der Inanspruchnahme der Krebsfrüherkennungsprogramme sowie
 - Fortschritte der Primärtherapie?

6. Sensitivität/Spezifität der benutzten Tests

- 6.1 Welche Ereignisse können vom Test prinzipiell entdeckt werden, welche entziehen sich der Entdeckbarkeit?
- 6.2 Sind Sensitivität und Spezifität bekannt und zufriedenstellend?
- 6.3 Auf welche Daten stützen sich die Angaben über Sensitivität bzw. Spezifität?
- 6.4 Entspricht das erwartete Spektrum der Krankheitsausprägungen und konkurrierenden Veränderungen den Voraussetzungen, unter denen der Test validiert wurde?
- 6.5 Wurde insbesondere berücksichtigt, daß sich in der Tumornachsorge aufgrund therapieinduzierter Veränderungen besondere die Spezifität mindernde differentialdiagnostische Probleme stellen können?

7. Prädiktiver Wert des positiven bzw. negativen Tests

- 7.1 Wie hoch ist der anzunehmende prädiktive Wert des Tests?
- 7.2 Wurde berücksichtigt, daß die Prävalenz des Zielereignisses von der Untersuchungsfrequenz im Nachsorgeprogramm abhängt?
- 7.3 Wurden Wechselwirkungen zwischen Prävalenz und Sensitivität bzw. Spezifität berücksichtigt, wie
 - Nachlassen der Aufmerksamkeit des Befunders bei niedriger Prävalenz und
 - allmähliche Verschiebung des Schwellenwertes aufgrund von Verunsicherung des Untersuchers bei wider Erwarten niedriger Trefferrate?

8. Zusammenwirken mehrerer Tests
 - 8.1 Stehen Referenzmethoden zur Verfügung?
 - 8.2 Welchen Stellenwert hat ein Test im Gesamtprogramm:
 - werden Tests simultan zur Hebung der Spezifität eingesetzt?
 - werden Tests simultan zur Hebung der Sensitivität eingesetzt?
 - werden befundabhängig weitere Tests erforderlich?
 - soll der Test bei positivem Ausgang wiederholt werden?
 - soll der Test bei negativem Ausgang wiederholt werden?
9. Beeinträchtigung von Sensitivität und Spezifität in der Nachsorge aufgrund von Abweichungen von definierten Standards
 - 9.1 Ist die zur Erreichung der erwarteten Präzision erforderliche Qualifikation bzw. Routine gegeben?
 - 9.2 Genügt der technische Standard den Anforderungen?
 - 9.3 Spielen weitere autonome Eigenschaften des Untersuchers eine Rolle, z.B. individuelle Ansichten über die Tragweite eines Befundes bzw. den Stellenwert der verfügbaren Therapie?
 - 9.4 Variiert die Testdurchführung, ist der Test hinreichend normiert?
 - 9.5 Wird die Indikationsstellung zur Anwendung bestimmter Verfahren einheitlich und adäquat gehandhabt?
 - 9.6 Ist das Verhalten bei mehrdeutigem Befund hinreichend standardisiert?
 - 9.7 Ist sichergestellt, daß erforderliche Untersuchungen oder die adäquate Befundung nicht durch Unter- bzw. Überkapazität beeinträchtigt werden?
 - 9.8 Ist sichergestellt, daß nicht weitere Faktoren, z.B. entfernter Wohnort des Patienten, die Befundung beeinträchtigen?
10. Unterstützung korrekter Befundung durch organisatorische Maßnahmen
 - 10.1 Wird die Kommunikation der an der Nachsorge Beteiligten bestmöglich unterstützt?
 - 10.2 Entspricht insbesondere der verwendete Nachsorgepaß den Anforderungen einer multizentrischen ambulanten Nachsorge?
 - 10.3 Werden relevante Informationen zum bisherigen Krankheitsverlauf für die Nachsorge nutzbar gemacht?
 - 10.4 Werden Vorinformationen über bisherige Tests verfügbar?

- 10.5 Werden Informationen über parallel durchgeführte Tests verfügbar?
- 10.6 Sind die Vorgaben des Nachsorgeprogramms verständlich, präzise und logisch schlüssig?
- 10.7 Werden zusätzlich Hintergrundinformationen über Symptome, Abklärungsstrategien und Häufigkeiten zur Verfügung gestellt?
- 10.8 Wird der Bezug des Nachsorgeprogramms zum Gesamtverlauf transparent, z.B. durch Manuale zu Diagnostik, Therapie und Nachsorge?
- 10.9 Werden Maßnahmen zur Qualitätskontrolle der erbrachten Leistungen durchgeführt

11. Kapazität

- 11.1 Ist das Versorgungssystem in der Lage, den jetzt und in Zukunft durch ein Nachsorgeprogramm induzierten Diagnostikbedarf zu befriedigen?
- 11.2 Kann der im vorliegenden Nachsorgeprogramm festgeschriebene Aufwand ohne Kapazitätsüberschreitung auf vergleichbare Problemstellungen, insbesondere weitere Tumordiagnosen, verallgemeinert werden?
- 11.3 Stehen die Kosten pro Tumorpatient in ausgewogenem Verhältnis zum Gesamtaufwand für die medizinische Versorgung?

12. Stabilität und Flexibilität

- 12.1 Stützt sich das Programm auf eine stabile Basisorganisation?
- 12.2 Ist das Programm flexibel gegen Modifikationen der Strategie?

13. Informationsstrukturen zur Messung der Effektivität von Tumornachsorge

- 13.1 Sind Informationsstrukturen gegeben, die die Evaluation des Nachsorgeprogramms im Rahmen der Breitenversorgung gestatten?
- 13.2 Existiert ein verlaufsorientiertes Tumorregister?

- 13.3 Können die Verlaufsdaten aus der ambulanten Nachsorge mit den Primärdaten des Tumorregisters bei Wahrung der Datenschutzerfordernisse verknüpft werden?

14. Durchführbarkeit von Fall-Kontroll-Studien

- 14.1 Ist die Klassifikation der im Einzelfall durchgeführten Nachsorge ('Exposition') definiert?

- 14.2 Wurden geeignete Zielkriterien des medizinischen Nutzens gewählt?
- 14.3 Ist das anhand des Zielkriteriums definierte Zielereignis in jedem Einzelfall entscheidbar?
- 14.4 Wurden die bekannten Effekte des Lead-Time bzw. Length Bias bedacht, insbesondere auch im Zusammenhang mit weiteren Zielkriterien, wie z.B. Überleben ab Diagnose einer Progression?
- 14.5 Ist die Rekrutierung von Fällen und Kontrollen auf der Basis des gewählten Kriteriums möglich und unabhängig von der laufenden Nachsorge durchführbar?
- 14.6 Sind Schichtkriterien definiert und verfügbar?
15. Dokumentation
- 15.1 Werden die in der Nachsorge erhobenen Befunde dokumentiert?
- 15.2 Ist der Inhalt der Routinedokumentation praxisverträglich?
- 15.3 Sind die erhobenen Daten geeignet
- zur Beschreibung des Krankheitsverlaufs,
 - als Basis für Fall-Kontroll-Studien,
 - als Beitrag zur Qualitätskontrolle der Primärtherapie,
 - als Basis für weitergehende Studien,
 - als Input für mathematische Modelle,
 - zur Quantifizierung des Versorgungsaufwands und
 - zur Analyse des Inanspruchnahmeverhaltens?
16. Inanspruchnahme
- 16.1 Wurde Vorsorge getroffen, daß High-Risk-Gruppen gezielt betreut werden können?
- 16.2 Ist sichergestellt, daß jede Klinik erfährt, welche ihrer Patienten regelmäßig in Nachsorge stehen und welche verstorben sind?
- 16.3 Existieren effiziente Informationswege von den Kliniken zu den niedergelassenen Ärzten, um Nichtteilnehmern gezielt das Versorgungsangebot nahebringen zu können?
-

5.3 Zur Umsetzbarkeit methodischer Forderungen

Die umfangreiche Anforderungsliste in Abschnitt 5.2 scheint dem praktischen Handeln enge Grenzen zu setzen. Insbesondere für umfangreiche Programme ergibt sich in der Tat ein Informationsbedarf, dem vor Aufnahme der Aktivität schwer entsprochen werden kann. Vorgeschlagen wird daher die stufenweise Einführung von Nachsorge, beginnend mit einem Basisprogramm, welches sich weit-

gehend auf klinische Untersuchung und Anamnese beschränkt (Tab. 5.3). Nachsorgeprogramm soll hierbei entsprechend der Definition in Abschnitt 2 als Versorgungsangebot an klinisch tumorfreie Patienten verstanden werden, für die keine unmittelbare Notwendigkeit zur ärztlichen Intervention besteht.

Beispielsweise in Bayern werden derzeit nur für einige Tumordiagnosen Nachsorgeprogramme angeboten. Die folgenden Vorschläge zielen daher nicht auf die Einengung notwendiger Versorgungsleistungen, sondern auf den stufenweisen, kontrollierten Ausbau von bisher noch nicht oder kaum erbrachten Leistungen ab. Als Beginn eines allmählichen Aufbaus von tumorspezifischen Nachsorgeprogrammen wird folgendes vorgeschlagen.

Tab. 5.3: Vorschlag für ein Basis-Nachsorgeprogramm zur Gewährleistung einer minimalen Tumornachsorge für alle Patienten innerhalb der ersten 5 (10) Jahre nach Primärtherapie

-
1. Jeder Tumorpatient hat in den ersten 5 bzw. 10 Jahren nach Abschluß seiner Primärbehandlung Anspruch auf eine vorgegebene Zahl von Nachsorgeuntersuchungen in definierten Zeitabständen.
 2. Die routinemäßig durchgeführten Nachsorgeuntersuchungen bestehen aus Zwischenanamnese, klinischer Untersuchung und gegebenenfalls Beratung des Patienten.
 3. Ergeben sich aus der Nachsorgeuntersuchung Indikationen für engmaschige Beobachtung, weitergehende Diagnostik oder Einleitung von Behandlungsmaßnahmen, so werden diese Maßnahmen entsprechend den Erfordernissen durchgeführt.
 4. Durchführung, Ergebnis und gegebenenfalls Konsequenzen der Nachsorgeuntersuchung sind zu dokumentieren. Die Daten sind ohne Nennung von Namen zentral und verlaufsbezogen auswertbar.
 5. Jedem Tumorpatienten wird nach Abschluß der Primärbehandlung ein Nachsorgekalender ausgehändigt, der die Kommunikation der gemeinsam die Nachsorge tragenden Ärzte unterstützt, und zugleich über eine eindeutige Nummer die Zuordnung bzw. Zusammenführung konsekutiver Dokumentationsbeiträge in der Zentrale ermöglicht.
-

Damit ist in den Grundzügen ein Programm skizziert, das für jeden Tumorpatienten eine minimale Betreuung sicherstellt, dabei angesichts der hohen Prävalenz den Rahmen des Finanzierbaren einhält und im Einzelfall die Durchführung weiterer, erkennbar notwendiger Maßnahmen nicht behindert.

Die medizinische Relevanz dieses Basisprogramms könnte angesichts der aufwendigeren Nachsorgeempfehlungen als gering eingeschätzt werden. Breit angelegte empirische Untersuchungen fehlen, ein interessantes Ergebnis aus England sei jedoch zitiert. An einer Gruppe von 180 regelmäßig nachgesorgten Colonkarzinom-Patienten wurden 71 Rezidive diagnostiziert, wobei 58% der Diagnosen außerhalb der Nachsorge gestellt wurden (8). Hieraus leiten die Autoren die Schlußfolgerung ab, daß Aufklärung des Patienten über relevante Symptome effektiver als eine zeit- und kostenintensive Nachsorge sein könnte. Im Sinne dieser Argumentation könnte die regelmäßige Zwischenanamnese beim Hausarzt nicht zuletzt als edukative Maßnahme zur Erhöhung der Sensibilität des Patienten gegenüber Veränderungen verstanden werden.

Die begleitende Dokumentation sollte sich auf die Kernfragen beschränken, um in der Breitenversorgung durchführbar zu sein. Sie soll die Wirkungen des Programms transparent machen, programmimmanent verfügbare Wirksamkeitskriterien erfassen, Erkenntnisse zum natürlichen Krankheitsverlauf liefern sowie die Durchführung von Spezialstudien an Untergruppen ermöglichen.

Das Konzept wird abgerundet durch die Verknüpfung der aus dem Basis-Nachsorgeprogramm angefallenen Verlaufsdaten mit den in einem Tumorregister gespeicherten Primärdaten, soweit solche verfügbar sind.

Für die Gegebenheiten in Südbayern wurde ein solches Datenflußkonzept, in dem Registerdaten aus dem Tumorregister München über die Nummer des Nachsorgepasses mit anonymisierten Daten aus der ambulanten Nachsorge verknüpft werden können, bereits in (26) und erstmals in (25) beschrieben.

Fußend auf der so gewonnenen Evidenz könnte dieses Basis-Nachsorgeprogramm stufenweise ausgebaut werden. Die Ausweitung der Aktivitäten über dieses Grundprogramm hinaus kann so an die Entwicklung der aus regulärer Nachsorge und begleitenden Studien wachsenden Evidenz angepaßt werden, während eine medizinische Basisbetreuung, an deren Zweckmäßigkeit bisher von keiner Stelle Zweifel geäußert wurden, von Anfang an gewährleistet wäre.

6. Modellrechnungen zur Tumornachsorge

Die Synopse der Abschnitte 3, 4 und 5 belegt, daß bei der Planung und Bewertung von Nachsorgeprogrammen stellenweise die aus methodischer Sicht geeigneten Ansätze und Vorgehensweisen nicht oder nur eingeschränkt anwendbar sind. Die in Abschnitt 5.2 vorgelegte Anforderungsliste reflektiert das Bemühen, eine weithin als notwendig angesehene Entwicklung unter Berücksichtigung von nicht veränderbaren Randbedingungen methodisch zu begleiten.

Im Rahmen der folgenden Modellrechnungen soll abschließend untersucht werden, inwieweit vorhandenes Datenmaterial, vorwiegend Registerdaten, zur Schätzung von Planungsgrößen geeignet ist.

Zur Abschätzung der mutmaßlichen Inanspruchnahme von Nachsorgeprogrammen wird in Abschnitt 6.1 die zukünftige Entwicklung der Prävalenz Tumorkranker extrapoliert. In Abschnitt 6.2 wird der Altersaufbau der derzeit im 5-Jahres-Intervall nach Primärbehandlung Befindlichen analysiert.

Am Beispiel Mammakarzinom wird anschließend (Abschn. 6.3) versucht, Metastasierungswahrscheinlichkeiten aus Registerdaten zu schätzen und einer praktizierten Nachsorgestrategie gegenüberzustellen. Durch diese Simulation werden die in Abschnitt 3 angestellten Überlegungen zur Übertragbarkeit anderweitig gewonnener Planungsgrößen auf die Situation in Nachsorgeprogrammen anhand eines praktischen Beispiels illustriert.

6.1 Zur zukünftigen Entwicklung der Prävalenz Tumorkranker

6.1.1 Extrapolation der Neuerkrankungsraten

Krebs ist vorwiegend eine Erkrankung des hohen und höchsten Lebensalters. Dies impliziert nicht, daß für jede einzelne Tumorklassifikation zwischen Alter und Inzidenz eine positive Korrelation besteht. Kumuliert über alle Lokalisationen weisen jedoch Inzidenz und Mortalitätsziffern der Krebserkrankungen weltweit eine positive Korrelation mit dem Alter auf.

Für die Mehrzahl der relevanten Fragestellungen ist diese Altersabhängigkeit als Störfaktor anzusehen. Insbesondere zur Gewinnung großräumig vergleichbarer populationsbezogener Inzidenzangaben ist deshalb eine Altersstandardisierung der beobachteten Raten erforderlich. Diese wird durchgeführt, indem die pro Altersklasse beobachteten Inzidenzen mit dem fiktiven Anteil dieser Altersklasse an einer Normpopulation gewichtet werden. Alterstandardisierte Inzidenzen quantifizieren die Auswirkungen der in der Bezugspopulation wirksamen Risiken auf die gewählte Normpopulation.

Die Größe des in Wirklichkeit durch Krebs induzierten Gesundheits- und Versorgungsproblems wird dagegen durch standardisierte Maßzahlen um so ungenauer approximiert, je stärker Normpopulation und Bezugspopulation sich im Aufbau unterscheiden. Vorzuziehen ist in diesem Fall die rohe, d.h. nicht standardisierte Inzidenz ('Crude Rate'). Ändert sich der Altersaufbau der Bevölkerung, so ändert sich auch die Crude Rate.

6.1.1.1 Daten und Annahmen

Im folgenden wird der Verlauf der Crude Rate für die erwartete Population der Bundesrepublik bis zum Jahre 2030 extrapoliert. Dies erfordert Daten über die altersspezifische Inzidenz von Tumorerkrankungen, Daten über die Altersstruktur der Bevölkerung sowie Annahmen über deren zukünftige Entwicklung. Inzidenzdaten liegen weder für die gesamte Bundesrepublik noch für Bayern vor. Als Basis der folgenden Modellrechnungen wurden die im Krebsregister des Saarlandes ermittelten Inzidenzen der Jahre 1979-81 (60) herangezogen.

Für eine Zunahme der altersspezifischen Krebsgefährdung besteht derzeit aufgrund von Analysen der Mortalitätsstatistiken kein Anhalt (37, 46). Zwar wird die These der Konstanz der Krebsgefährdung teils kontrovers diskutiert, Spekulationen über ein zukünftiges ökologisch bedingtes Anwachsen des Krebsrisikos sind jedoch zu vage, als daß sie im Rahmen einer Extrapolation berücksichtigt werden könnten. Für die folgende Modellrechnung wird deshalb lediglich angenommen, daß die altersspezifische Gefährdung nicht sinkt. Sollte sie tatsächlich ansteigen, so wären die Schätzwerte entsprechend nach oben zu korrigieren.

Auch die Extrapolation des Altersaufbaus einer Bevölkerung ist zahlreichen Imponderabilien ausgesetzt, insbesondere angesichts der Situation der Bevölkerung der Bundesrepublik, die derzeit die weltweit niedrigste Geburtenrate aufweist. Während für - aus der Sicht der Bevölkerungswissenschaft - kurze Zeiträume von unter 10 Jahren der Altersaufbau nur träge auf Änderungen der Nettoreproduktionsrate anspricht, müssen für mittelfristige Zeiträume infolge alternativer, jeweils plausibel begründbarer Annahmen breit gefächerte Schätzungen hingenommen werden.

Die im folgenden benutzten aktuellen und extrapolierten Bevölkerungszahlen sind der Monographie (6) entnommen.

6.1.1.2 Methodik der Extrapolation

Unter der Annahme des konstanten Krebsrisikos erfolgt die Extrapolation der Crude Rate, indem die aktuellen altersspezifischen Inzidenzen mit dem Anteil der jeweiligen Altersklasse an der prognostizierten Population gewichtet werden. Rechnerisch entspricht dies der direkten Standardisierung mit dem prognostizierten Aufbau der bundesdeutschen Bevölkerung. Für den kritischen Anteil der über 60-jährigen an der Gesamtbevölkerung im Jahre 2030 werden in (6) Werte zwischen 24,5% und 32% prognostiziert (Abb.6.1). Dies entspricht, bezogen auf den Anteil von 18,2% im Jahre 1980, einer Steigerung zwischen 35 und 76%. Damit ist auch die Extrapolation der Crude Rate entsprechenden Schwankungen ausgesetzt.

Im folgenden wird jeweils der Mittelwert der unter den Alternativen AI, AII und AIII (Abb. 6.1) errechneten Extreme zugrundegelegt.

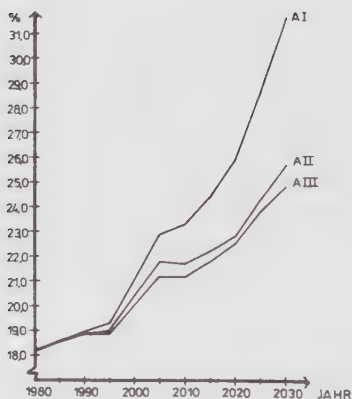


Abb. 6.1: Entwicklung des Anteils der über 60-jährigen in Prozent an der Gesamtbevölkerung unter alternativen Annahmen über die Nettoreproduktionsrate

A I: Beibehaltung der Fertilitätsrate des Jahres 1979 über 50 Jahre

A II: Allmählicher Anstieg der Fertilitätsrate bis zum Jahr 2000, danach konstante Verhältnisse

AIII: Zyklisches Verhalten der Fertilitätsrate

(entnommen aus (6, S. 46)

Im folgenden wird jeweils der Mittelwert der unter den Alternativen AI, AII und AIII (Abb. 6.1) errechneten Extreme zugrundegelegt.

Die Inzidenzen des Saarländischen Krebsregisters werden auf breitere, mit der Prognose der Bevölkerungsentwicklung kongruente Altersklassen umgerechnet. Zusammengefaßt werden die Altersgruppen 0-19, 20-39, 40-59 sowie 60 Jahre und älter. Für die Klassen bis unter 60 Jahren werden die zugehörigen Inzidenzen anhand des arithmetischen Mittels der Inzidenzen der korrespondierenden 5-Jahres-Intervalle geschätzt.

Für die höchste Altersklasse wird die durchschnittliche Inzidenz durch die Beziehung

durchschnittl. Inzidenz >60 * Anteil der Altersklasse >60

$$= \text{Crude Rate} - \sum_{i=1}^3 \text{durchschnittl. Inzidenz}_i \cdot \text{Anteil der Altersklasse}_i$$

geschätzt. Eine Verwendung des arithmetischen Mittels hätte angesichts des raschen Anstiegs der Inzidenz in den höchsten Altersgruppen bei gleichzeitiger Abnahme der Besetzungszahlen zu Verzerrungen geführt.

6.1.1.3 Ergebnisse der Extrapolation

Die für die Altersklassen 0-19, 20-39, 40-59 sowie 60 und älter gemittelten Inzidenzen sind in Tabelle 6.1 aufgeführt. Tab. 6.2 beschreibt die Entwicklung der Crude Rate bis zum Jahr 2030 auf der Basis dieser Werte.

Tab. 6.1: Ermittlung von zur Extrapolation geeigneten Inzidenzen aus den Daten des Krebsregisters des Saarlandes 1979-81

Altersklasse	gemittelte Inzidenz männlich	gemittelte Inzidenz weiblich	% an der Gesamt- bevölkerung
0 - 19	13.85	9.63	26.5
20 - 39	51.93	63.30	29.8
40 - 59	445.48	403.95	25.5
60 und älter	1723.90	1554.42	18.2
Crude Rate	446.50	407.30	100.0

Tab. 6.2: Extrapolation der Rate der Neuerkrankungen an malignen Tumoren (Crude Rate) in der Bundesrepublik für den Zeitraum 1980 - 2030 bei altersspezifisch konstanter Krebsgefährdung. Altersaufbau nach (6); Inzidenzen nach (60) und Tab. 6.1

Anteil der Altersgruppe*)	1980	2000	2030
0 - 19	26.5 %	22.5 %	21.2 %
20 - 39	29.8 %	29.8 %	26.1 %
40 - 59	25.5 %	27.6 %	24.6 %
60 und älter	18.2 %	20.1 %	28.1 %
Crude Rate männl.	446.5	488.0	610.5
weibl.	407.3	444.1	554.7
Steigerung		9 %	36 %

*) Die Anteile wurden geschätzt, indem jeweils zwischen den unter AI, AII bzw. AIII (vgl. Abb. 6.1) ermittelten Extremen gemittelt wurde.

Tab. 6.3: Entwicklung der nach Tab. 6.2 jährlich zu erwartenden Neuerkrankungen bis zum Jahr 2030 in absoluten Zahlen

	1980	2000	2030
Einwohner (in 1000) geschätzt*)	61.145,6	57.991,5	48.150,6
Crude Rate (1/2 . männl. + 1/2 weibl.)	426,9	466,1	582,6
Neuerkrankungen (in 1000)	261	270	281

*) vgl. Fußnote zu Tab. 6.2

Aus der Schätzung wird erkennbar, daß der tumorbedingte Versorgungsaufwand bis zum Ende des Jahrtausends nur geringfügig ansteigen dürfte. Danach werden die besonders krebgefährdeten Altersklassen jedoch stärker an Gewicht zunehmen. Mittelfristig ist daher mit einer erheblichen Steigerung der Zahl der Neuerkrankun-

gen pro 100.000 der Bevölkerung in der jeweiligen Zusammensetzung zu rechnen. In absoluten Zahlen wäre trotz eines Rückgangs der Bevölkerung um etwa 13 Mio. im Jahre 2030 etwa dieselbe Zahl von Neuerkrankungen zu erwarten wie heute, wie Tabelle 6.3 zeigt.

Ein ungünstigeres Bild entsteht, wenn die Neuerkrankungsraten auf die Gruppe der Erwerbsfähigen bezogen werden. Nach (6) ist bis 2030 mit einem Rückgang der Erwerbsfähigenquote von derzeit 55 auf etwa 50% zu rechnen. Im gleichen Zeitraum steigt die Crude Rate von 427 auf etwa 583 an. Zusammen ergibt sich für die Rate der Neuerkrankungen auf 100.000 Erwerbsfähige eine Steigerung um 50% von derzeit 776 auf 1165 im Jahre 2030.

6.1.2 Weitere für die Entwicklung der Prävalenz Tumorkranker bestimmende Faktoren

Die Prävalenz ergibt sich aus dem Produkt von Inzidenz und durchschnittlicher Dauer. Letztere ist bei chronischen Krankheiten wie Krebs mit der Überlebensdauer gleichzusetzen und wird durch zahlreiche prognoserelevante Kriterien, wie

- relative Häufigkeit der einzelnen Tumorarten,
- Wirksamkeit der tumorspezifischen Therapie und
- Teilnahme an Vorsorgeuntersuchungen

beeinflusst.

Bekanntlich ist die Art einer Tumorerkrankung, charakterisiert durch Lokalisation und/oder Histologie, von überragender Bedeutung für die Überlebenschancen des Patienten. Zumindest für frühe Stadien variieren die 5-Jahres-Überlebensraten von nahezu 100%, wie z.B. beim Cervixkarzinom, bis nahezu 0%, z.B. beim kleinzelligen Bronchialkarzinom. Aus Längsschnittuntersuchungen an Mortalitätsdaten sind großräumig reproduzierbare und lang anhaltende Trends der Zunahme bzw. Abnahme einzelner Tumorarten beschrieben (z.B. 37, 66).

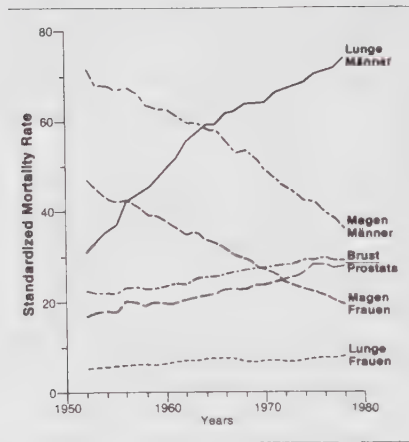


Abb. 6.2: Zeitlicher Trend einiger Krebsformen in der Bundesrepublik Deutschland in den letzten 30 Jahren, entnommen aus (66).

Eine Extrapolation solcher Trends erscheint problematisch. Einerseits basieren diese auf Mortalitätsdaten, die teils erheblich von den Inzidenzen abweichen. Diese Abweichung hängt, abgesehen von Registrierungslücken und Fehlklassifikationen, von der für eine Tumorart anzunehmenden Quote von Dauerheilungen ab. Je größer die Heilungschancen werden, desto geringer der Anteil der Erkrankten, der sich in der Mortalitätsstatistik manifestiert. Dies ist einer der Gründe für die weithin gestellte Forderung nach weiteren Inzidenzregistern. Ob Änderungen der Mortalitätsstatistik in erster Linie auf therapeutische Fortschritte oder auf Änderungen der tumorspezifischen Inzidenzen zurückzuführen sind, bleibt häufig unbeantwortet.

Andererseits ist die Überlebenswahrscheinlichkeit bedingt durch die verfügbare Therapie. Die Extrapolation der relativen Häufigkeit einzelner Tumorarten erhielte erst dann wirklichen Sinn, wenn auch die Wirksamkeit der verfügbaren tumorspezifischen Behandlungsmöglichkeiten prognostiziert würde. Dies wirft die Frage auf, wie Innovationen vorhergesagt werden können. Es ist wohl gerechtfertigt, von einem stetigen Trend zur Verbesserung der Tumorthherapie insgesamt auszugehen.

Es ist aber sicher nicht gerechtfertigt, eine grundsätzliche Lösung des Problems für die nähere Zukunft in Aussicht zu stellen oder zu erwarten. Die bisher verzeichneten Fortschritte in der Behandlung Tumorkrankter sind als Summe tumorspezifischer und nicht spezifischer Innovationen entstanden. Vieles spricht dafür, daß sich diese Entwicklung in ähnlicher Weise fortsetzen wird.

Das Tumorstadium bei Diagnosestellung ist ein weiterer bestimmender Faktor für die Prognose. Es wird u.a. vom Ausmaß der Teilnahme an Krebsvorsorgeprogrammen verändert. Ob Krebsvorsorge zu einem echten Gewinn an Lebensjahren führt, ist nicht abschließend beantwortet. Gesichert ist aber ein anderer Effekt der Früherkennung, nämlich die Verlängerung der Überlebenszeit ab Diagnosestellung, die durch Vorverlegung des Diagnosezeitpunktes entsteht und insofern unabhängig vom Gewinn an Lebensjahren ist. Aus der Sicht der Bewertung von Screeningprogrammen wird dieser Effekt als Bias bezeichnet, der sogenannte Lead Time Bias. Aufgrund dieses Effekts hängt die Prävalenz der in Nachsorge Stehenden u.a. von der Inanspruchnahme der Krebsfrüherkennungsprogramme ab.

Welche Annahme für die zukünftige Teilnahme an der Krebsfrüherkennung ist angemessen? Wird der derzeitig fallende Trend sich bis zur Einstellung fortsetzen, wird es zur Stagnation kommen, ist eine dauerhafte Trendwende zu erwarten oder ist die Teilnahme ein zyklischer Prozeß?

Weitere Imponderabilien der zukünftigen Lebenserwartung Tumorkrankter könnten genannt werden, wobei der wissenschaftliche Fortschritt wohl die entscheidende Rolle spielt. Angesichts dieser Vielfalt schwer wägbarer, synergistisch oder antagonistisch wirkender Faktoren sind höchstens grobe globale Schätzungen für die zukünftige Entwicklung der Überlebenszeiten möglich.

In den Vereinigten Staaten hat sich nach Angaben des NCI die 5-Jahres-Überlebensrate aller Tumorpatienten insgesamt während der letzten 20 Jahre nach oben bewegt. Während die Wahrscheinlichkeit, 5 Jahre zu überleben für Patienten, die in den Jahren 1960-63 diagnostiziert wurden, noch 38% betrug, ergab sich für die Zeit von 1970-73 ein Wert von 42%, während zwischen 1973 und 1979 48% erreicht waren (zit. nach (9)).

Damit wurde die Quote der frühzeitig Verstorbenen in 20 Jahren um über 16% reduziert. Bei Fortsetzung dieses Trends würde sich bis zum Jahr 2030 eine 5-Jahres-Überlebensrate von ca. 65% ergeben, gemessen an derzeit 48% eine Steigerung von 35%. Diese Ziffer mag als globales Maß dafür angesehen werden, wie sich die Überlebenswahrscheinlichkeit unter dem Einfluß von Verschiebungen im Spektrum der Tumorarten, Verbesserungen der Therapie und einem sich möglicherweise ändernden Gesundheitsbewußtsein der Bevölkerung entwickeln könnte. Über die Verwendung und Interpretation dieser aus amerikanischen Verhältnissen ermittelten Werte hinaus sind keine weitergehenden Vorhersagen möglich, da in der Bundesrepublik keine repräsentativen Verlaufsregister existieren.

6.1.3 Diskussion der Prognosen

Für die Extrapolation der zu erwartenden 5-Jahres-Überlebensraten trat die Situation auf, daß bereits die unmittelbar als relevant erkennbaren Einflußgrößen in ihrer zukünftigen Entwicklung kaum prognostizierbar sind. Daher wurde auf detaillierte Simulation verzichtet.

Dagegen wurde für die Entwicklung der Inzidenz unter der Annahme der altersspezifischen Konstanz der Krebsgefährdung eine Extrapolation über 50 Jahre versucht. Wegen der Mittelwertbildung aus alternativen Bevölkerungsprognosen und der Zusammenfassung zu relativ breiten Altersklassen hat auch diese Modellrechnung nur überschlägigen Charakter.

Andererseits existieren jedoch weitere Faktoren, die die Präzision der Schätzungen limitieren, so daß eine genauere Extrapolation nach Altersklassen und Geschlecht wohl nur scheinbar bessere Resultate erbracht hätte. Limitierend für die Präzision ist insbesondere die Tatsache, daß die im Saarland nachgewiesenen Inzidenzen allein nicht als repräsentativ für die Bevölkerung des Bundesgebiets gelten können und auch durch Synopse beider derzeit verfügbarer Inzidenzregister keine Repräsentativität für die Bundesrepublik erreichbar wäre.

Unabhängig von diesen Limitierungen stellt sich die Frage, inwieweit derartige Extrapolationen des Mengengerüsts zur Planung oder Begründung langfristiger Versorgungskonzepte dienlich sind. Dies gilt in besonderer Weise für fakultative Leistungen wie Tumornachsorge, für deren Inanspruchnahme die Kenntnis des Mengengerüsts allein keine Prognose gestattet, wie das Beispiel der Krebsfrüherkennungsprogramme in der Bundesrepublik zeigt.

Ein großer Teil des prognostizierbaren Zuwachses an potentiell nachzusorgenden Tumorpatienten geht auf Umschichtungen der Altersstruktur zurück, wobei für die Gruppe der alten und sehr alten Patienten besondere Überlegungen anzustellen sind. Nicht nur wird ein volles Ausreizen der jeweiligen therapeutischen Möglichkeiten an sehr alten Patienten vielfach abgelehnt. Auch der Bedarf an Tumornachsorge geht in dieser Altersklasse zurück, denn mit zunehmendem Alter nimmt die Mobilität und die Inanspruchnahme von fakultativen Versorgungsleistungen ab, und gleichzeitig wächst der Anteil derer, die ohnehin wegen verschiedener Leiden in ärztlicher Dauertherapie stehen. Für solche Patienten bedeutet Tumornachsorge einen kaum meßbaren Mehraufwand, da sie stark mit anderen Problemen verflochten ist.

Dem steht allerdings der Trend gegenüber, mit zunehmender Lebenserwartung bis in immer höheres Alter hinein ein aktives, 'lebenswertes' Leben zu führen, für das sich auch ein gewisser Einsatz 'rentiert', einschließlich der Teilnahme an Tumornachsorgeprogrammen.

Neben dem altersunabhängigen Gesundheitsbewußsein ist die Angebotsstruktur zukünftiger Nachsorgeprogramme selbst entscheidend für die Nachfrage (5). In jedem Fall sollte die Ausschöpfung der Zielgruppe zu den Zielen zukünftiger Nachsorgeprogramme zählen. Insofern sind die durchgeführten Schätzungen zum Mengengerüst durchaus informativ.

Zusammenfassend läßt sich für die kommenden 50 Jahre ein erheblicher Zuwachs der potentiellen Nachfrage nach Nachsorgeleistungen prognostizieren:

1. Für die tatsächliche Inzidenz (Crude Rate) ist ein Zuwachs von 30-40% bis zum Jahr 2030 aufgrund der Umschichtung der Alterspyramide wahrscheinlich. Ein etwaiges Anwachsen des Krebsrisikos ist hierbei nicht berücksichtigt.
2. Im gleichen Zeitraum könnte sich die über alle Tumorarten gemittelte 5-Jahres-Überlebensrate um etwa 30-40% erhöhen. Diese Schätzung basiert unter Vernachlässigung aller Details auf Extrapolationen der Therapieergebnisse in den letzten 20 Jahren in den USA.
3. Die Erhöhung der Prävalenz korrespondiert mit einem Rückgang der Erwerbsfähigenquote von 55% auf 50%, so daß die Finanzierung von mehr Behandlungsbedarf auf weniger Schultern lasten wird.

Gleichzeitig lehrt die Erfahrung, daß der wissenschaftliche Fortschritt bisher kontinuierlich zur Änderung von Standards und Hebung von Ansprüchen und damit zur Kostensteigerung geführt hat. Bei der Initiierung kostenwirksamer Programme in der Tumornachsorge scheint aufgrund dieser Ergebnisse ein vorsichtiges Vorgehen empfehlenswert.

6.2 Zur aktuellen Prävalenz Tumorkranker in Bayern

Die Häufigkeit der Krebserkrankung setzt Sachzwänge in zwei Richtungen. Erstens gewinnt Abhilfe Priorität, auch Abhilfe durch nicht-primäre Prävention, u.a. Nachsorge, und zweitens setzt gerade die Häufigkeit dem Handeln Schranken, da Ressourcen nicht beliebig vermehrbar sind. Prognosen zur Entwicklung der Crude Rate wurden in Abschn. 6.1 erstellt. Aus den folgenden altersspezifischen Prävalenzschätzungen können zusätzlich Orientierungshilfen zur voraussichtlichen Inanspruchnahme angebotener Leistungen abgeleitet werden. Geschätzt wird die derzeitige altersspezifische Prävalenz Tumorkranker in Bayern, geschichtet nach der Länge des zurückliegenden Krankheitsabschnitts.

6.2.1 Daten

Als Basiszeitraum für ein Nachsorgeprogramm wird im folgenden das Intervall von 5 Jahren nach Primärtherapie angesehen. Zur Ermittlung der Prävalenz an Tumorkranken, deren Primärbehandlung höchstens 5 Jahre zurückliegt, wurden folgende Daten herangezogen.

Alters- und geschlechtsspezifische Inzidenzen:

Benutzt wurden die vom Saarländischen Krebsregister für 1979 bis 1981 nachgewiesenen Inzidenzen (60).

Altersstruktur der Zielpopulation:

Die Altersstruktur der bayerischen Bevölkerung für beide Geschlechter kann aus den Unterlagen des Bayerischen Statistischen Landesamtes (4) direkt entnommen werden.

Sterberisiko der Zielpopulation:

Die Wahrscheinlichkeit, im Alter x innerhalb des folgenden Jahres zu sterben, ist in den Sterbetafeln für die bundesdeutsche Bevölkerung (61) verfügbar.

Bedingte Wahrscheinlichkeit, tumorunabhängig zu sterben:

Aus der amtlichen Mortalitätsstatistik 1979-81 für Südbayern (3) wurden hierzu für jede Altersklasse und beide Geschlechter die Anteile der tumorunabhängig Verstorbenen an der Gesamtzahl der Verstorbenen in dieser Klasse berechnet.

Wahrscheinlichkeit, innerhalb von 5 Jahren nach Diagnosestellung an einem Tumor zu sterben (5-Jahres-Raten):

Diese Daten wurden überwiegend einem Lehrbuch der inneren Medizin (15) entnommen, um die Variation publizierter Einzelergebnisse auszugleichen (Tab. 6.4, S. 134).

Anhand dieser Quellen wurde für beide Geschlechter und jede Altersklasse die durchschnittliche 5-Jahres-Überlebensrate geschätzt. Hierzu wurden die Wahrscheinlichkeiten, innerhalb von 5 Jahren

- tumorabhängig oder
- tumorunabhängig

zu sterben, in geeigneter Weise verknüpft. Dies wird nachfolgend beschrieben.

Tab. 6.4: 5-Jahres-Überlebensraten (15) zur Schätzung der Prävalenz Tumorkrankter in Bayern.
Wiedergegeben sind die häufigeren Diagnosen

ICD	Bezeichnung	5-Jahres-Rate (%)
140-149		
160-161	HNO-Bereich	50
151	Magen	12
153-4	Dickdarm	38
162	Lunge	7
175	Mamma	65*)
185	Prostata	51
180	Cervix/Corpus uteri	60
183	Ovar	22
188	Harnblase	56
189.0	Niere	36

*) Die Prognose des Mammarkarzinoms wurde aus (56) ergänzt.

6.2.2 Methodik der Prävalenzschätzung

Die aus der Bevölkerungsstatistik zu entnehmenden Sterbewahrscheinlichkeiten für jede Altersklasse und beide Geschlechter beinhalten zwar sowohl die Tumorkrankheit als auch alle anderen Risiken, beziehen sich aber auf zufällig ausgewählte Individuen aus der Gesamtbevölkerung.

Die Teilmenge der tumortragenden bayerischen Bevölkerung weist ein höheres Risiko auf, tumorbedingt zu sterben. Daher muß die Wahrscheinlichkeit, dem bereits vorhandenen Tumorleiden zu erliegen, aus anderen Quellen ermittelt werden. Jedoch wird angenommen, daß diese Gruppe bezüglich des Risikos, tumorunabhängig zu sterben, als Zufallsstichprobe der Gesamtbevölkerung gleichen Alters und Geschlechts betrachtet werden darf. Unter Berücksichtigung beider Klassen von Todesursachen ergibt sich die alters- und geschlechtsspezifische 5-Jahres-Überlebensrate der tumortragenden Bevölkerung l_5 (..) wie folgt:

$$1_5(\text{Alter, Geschlecht}) = 1 - u_5(\text{Alter, Geschlecht})$$

$$- a_5(\text{Alter, Geschlecht, kein
tumorunabhängiges Ableben})$$

wobei $u_5(\dots)$ bzw. $a_5(\dots)$ die Wahrscheinlichkeiten bezeichnen, innerhalb von 5 Jahren unter den in der Klammer genannten Bedingungen tumorunabhängig (u_5 , bzw. tumorabhängig (a_5) zu sterben. $u_5(\dots)$ und $a_5(\dots)$ sind also definitionsgemäß disjunkt.

Für das weitere Vorgehen ist es von Wichtigkeit festzuhalten, daß den bedingten Wahrscheinlichkeiten $a_5(\dots)$ gerade die publizierten 5-Jahres-Überlebensraten entsprechen, wenn diese nach der Sterbetafeltechnik unter Elimination des tumorunabhängigen Ablebens als konkurrierendes Risiko ermittelt werden.

Zur quantitativen Bestimmung des alters- und geschlechtsspezifischen tumorbedingten Absterbens a_5 wurden die 5-Jahres-Raten der einzelnen Tumorarten mit der jeweiligen Inzidenz gewichtet und gemittelt,

$$a_5(\text{Alter, Geschlecht}) = \frac{1}{\text{Inzidenz}(\text{Alter, Geschlecht})} \cdot$$

$$\sum_{\text{Tumordiagnosen}_i} a_5(\text{Alter, Geschlecht, Tumor}_i) \cdot \text{Inzidenz}(\text{Alter, Geschlecht, Tumor}_i)$$

Für die Ermittlung der Wahrscheinlichkeit des tumorunabhängigen Absterbens u_5 wurden die Sterbewahrscheinlichkeiten für die deutsche Bevölkerung und die aus der bayerischen Mortalitätsstatistik zu entnehmenden relativen Häufigkeiten der tumorbedingten Sterbefälle verknüpft. Die Wahrscheinlichkeit, innerhalb eines Jahres tumorunabhängig zu sterben, $u_1(\dots)$, wurde wie folgt zerlegt:

$$u_1(\dots) = p(\text{Tod tumorunabhängig} | \text{Tod tritt ein}) \cdot p(\text{Tod tritt ein}).$$

Der erste Term wurde nun durch die relative Häufigkeit der nicht tumorbedingten Sterbefälle in der Mortalität der jeweils nächstfolgenden Altersklasse, also der Altersklasse, in die ein Individuum im Laufe der 5-Jahres-Periode hineinwächst, geschätzt.

Der zweite Faktor, die Wahrscheinlichkeit, innerhalb eines Jahres zu sterben (nicht conditional) ist in den Sterbetafeln für die bundesdeutsche Bevölkerung in Intervallschritten von 5 Jahren tabelliert. Benutzt wurde der dem Ende der jeweiligen Altersklasse entsprechende Tabellenwert. Die Rate u_5 ergibt sich nun unter der Unabhängigkeitsvoraussetzung analog zur Berechnung einer Survival-Funktion in Sterbetafeln durch

$$u_5(\text{Alter}, \text{Geschlecht}) = 1 - (1 - u_1)^5.$$

Nach den so ermittelten Größen a_5 und u_5 wird das 5-Jahres-Nettoüberleben $l_5(\dots)$ ermittelt,

$$l_5(\text{Alter}, \text{Geschlecht}) = 1 - a_5(\text{Alter}, \text{Geschlecht}) - u_5(\text{Alter}, \text{Geschlecht}).$$

Hieraus ergibt sich unter der Annahme exponentialverteilter Überlebensdauern, d.h.

$$S(t) = \exp(-\lambda t),$$

der Parameter $\lambda(\text{Alter}, \text{Geschlecht})$ aus der Beziehung

$$\lambda(\text{Alter}, \text{Geschlecht}) = (1/60 \text{ Mon.}) \cdot -\ln(l_5(\text{Alter}, \text{Geschlecht}))$$

6.2.3 Ergebnisse der Prävalenzschätzung

Die folgenden Tabellen 6.5 bis 6.10 geben die Resultate der Prävalenzschätzung wieder. Die höchste Altersklasse 85+ wurde wegen geringer Relevanz für die Nachsorge nicht in die Modellrechnung einbezogen. Auf die Altersklassen bis unter 30 Jahren entfallen insgesamt nur 2% der Fälle. Diese wurden in der Modellrechnung berücksichtigt, jedoch unterbleibt die Wiedergabe der klassenspezifischen Werte für diese Altersgruppe aus Raumgründen. In der Spaltenpalte sind dagegen alle Werte enthalten, so daß die Summe sich jeweils auf alle Altersklassen von 0 bis unter 85 Jahren bezieht.

Aus der ersten Zeile der Tabellen 6.5 bis 6.7 entnimmt man die Neuerkrankungen maligner Tumoren für Männer, Frauen und beide Geschlechter gemeinsam in jeder Altersklasse und insgesamt. Wegen des Fehlens der höchsten Altersklasse liegt diese Zahl unterhalb der Crude Rate, die sich bei der Anwendung der saarländischen Inzidenzen auf die bayerische Bevölkerung ergeben würde. Zu deren

Tab. 6.5: Ergebnisse der Prävalenzschätzung (Bayern) für das männliche Geschlecht: Prozentangaben vertikal

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
BEI DIAGNOSESTELLUNG	22644	294	258	806	1003	1779	1483	2279	3062	4519	3365	2147
NACH 0 MONATEN FOLLOW-UP	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NACH 3 MONATEN FOLLOW-UP	92.54	93.24	91.96	96.10	94.12	92.74	93.51	93.24	92.46	92.48	91.67	90.08
NACH 6 MONATEN FOLLOW-UP	85.64	91.16	93.91	99.05	96.41	96.07	97.04	96.97	95.43	95.53	93.67	91.14
NACH 9 MONATEN FOLLOW-UP	79.27	87.37	86.75	94.09	93.15	93.51	94.20	94.09	92.47	92.09	90.55	87.09
NACH 12 MONATEN FOLLOW-UP	71.54	82.39	81.79	90.14	90.17	90.72	91.57	91.40	90.29	90.14	88.01	85.81
NACH 15 MONATEN FOLLOW-UP	64.26	78.33	76.36	84.33	82.70	83.44	84.44	84.89	83.46	83.09	80.18	77.24
NACH 18 MONATEN FOLLOW-UP	57.53	71.17	71.14	80.03	78.27	78.84	80.04	80.75	79.18	78.37	75.20	72.84
NACH 21 MONATEN FOLLOW-UP	51.13	64.17	62.75	72.96	71.12	71.77	73.17	73.49	71.89	71.69	68.00	65.72
NACH 24 MONATEN FOLLOW-UP	45.27	58.27	57.29	68.21	66.34	66.34	67.51	67.68	66.17	66.17	62.51	60.28
NACH 27 MONATEN FOLLOW-UP	40.10	52.88	51.95	63.86	61.94	61.97	63.14	63.27	61.67	61.67	58.05	55.79
NACH 30 MONATEN FOLLOW-UP	35.84	47.63	46.65	58.58	56.76	56.55	57.64	57.77	56.09	56.09	52.41	50.14
NACH 33 MONATEN FOLLOW-UP	32.21	43.52	42.54	54.40	52.57	52.36	53.41	53.54	51.81	51.81	48.13	45.86
NACH 36 MONATEN FOLLOW-UP	29.51	40.32	39.34	50.25	48.42	48.21	49.26	49.39	47.66	47.66	44.00	41.73
NACH 39 MONATEN FOLLOW-UP	25.40	36.54	35.56	46.10	44.27	44.06	45.11	45.24	43.51	43.51	40.00	37.73
NACH 42 MONATEN FOLLOW-UP	21.07	32.46	31.48	41.94	40.11	40.04	41.09	41.22	39.49	39.49	36.00	33.74

5-JAHRESRATE DER NEUENTDECKUNG	0.99	0.99	0.98	0.98	0.98	0.97	0.96	0.96	0.97	0.97	0.95	0.94
WIE OBER, TUMOR AUSG. **)	0.99	0.99	0.98	0.98	0.98	0.97	0.96	0.96	0.97	0.97	0.95	0.94
TUMOR-FREIHEIT 5-JAHRES-RATE **)	0.40	0.37	0.32	0.40	0.37	0.37	0.37	0.37	0.37	0.37	0.35	0.33

*) WIE WURDE DIE BEVÖLKERUNG ABSTERBEN, WENN TUMORERKRANKUNGEN ALS TODESURSACHE ELIMINIERT WÄREN

**) WIE WURDEN TUMORERKRANKT ABSTERBEN, WENN ANDERE ERKRANKUNGEN ALS TODESURSACHE ELIMINIERT WÄREN

Tab. 6.6: Ergebnisse der Prävalenzschätzung (Bayern) für das weibliche Geschlecht: Prozentangaben vertikal

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
BEI DIAGNOSESTELLUNG	27392	751	554	792	1149	1502	1981	2340	2752	3380	3205	2351
NACH 0 MONATEN FOLLOW-UP	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NACH 3 MONATEN FOLLOW-UP	94.54	96.17	96.61	98.97	98.34	98.27	98.86	98.48	98.77	98.17	97.97	97.37
NACH 6 MONATEN FOLLOW-UP	87.41	92.72	93.22	96.07	95.72	95.74	96.92	96.54	96.79	96.07	95.37	94.50
NACH 9 MONATEN FOLLOW-UP	84.59	89.27	89.83	91.29	90.29	89.26	91.14	90.61	90.07	89.49	88.25	86.31
NACH 12 MONATEN FOLLOW-UP	80.34	85.82	86.72	88.51	87.98	86.72	88.50	87.80	87.00	86.54	84.57	82.76
NACH 15 MONATEN FOLLOW-UP	74.52	81.61	82.49	84.97	83.79	82.71	84.51	83.80	83.00	82.57	80.44	78.55
NACH 18 MONATEN FOLLOW-UP	69.17	77.78	78.11	81.57	79.79	78.76	80.57	79.86	79.00	78.57	76.44	74.55
NACH 21 MONATEN FOLLOW-UP	64.37	71.95	71.14	78.61	76.95	75.97	77.65	76.94	76.00	75.57	73.44	71.55
NACH 24 MONATEN FOLLOW-UP	59.91	70.11	71.67	75.75	74.00	73.07	74.54	73.84	72.90	72.47	70.34	68.45
NACH 27 MONATEN FOLLOW-UP	55.30	66.57	68.04	72.22	70.61	69.61	71.36	70.66	69.72	69.29	67.16	65.27
NACH 30 MONATEN FOLLOW-UP	52.70	63.50	64.97	69.37	67.59	66.55	68.30	67.60	66.66	66.23	64.10	62.21
NACH 33 MONATEN FOLLOW-UP	48.80	59.50	60.45	65.28	63.49	62.45	64.19	63.49	62.55	62.12	60.00	58.11
NACH 36 MONATEN FOLLOW-UP	44.17	54.41	55.21	61.16	59.36	58.32	60.06	59.36	58.42	57.99	55.86	53.97
NACH 39 MONATEN FOLLOW-UP	39.75	50.57	51.26	57.83	56.01	55.03	56.78	56.08	55.14	54.71	52.58	50.69
NACH 42 MONATEN FOLLOW-UP	34.34	46.74	48.59	54.62	52.77	51.73	53.41	52.71	51.77	51.34	49.21	47.30

5-JAHRESRATE DER NEUENTDECKUNG	0.99	0.99	0.99	0.98	0.97	0.96	0.96	0.93	0.92	0.93	0.95	0.98
WIE OBER, TUMOR AUSG. **)	1.00	1.00	1.00	0.99	0.98	0.98	0.98	0.95	0.94	0.93	0.93	0.92
TUMOR-FREIHEIT 5-JAHRES-RATE **)	0.47	0.49	0.55	0.47	0.48	0.44	0.43	0.37	0.36	0.33	0.32	0.32

*) WIE WURDE DIE BEVÖLKERUNG ABSTERBEN, WENN TUMORERKRANKUNGEN ALS TODESURSACHE ELIMINIERT WÄREN

**) WIE WURDEN TUMORERKRANKT ABSTERBEN, WENN ANDERE ERKRANKUNGEN ALS TODESURSACHE ELIMINIERT WÄREN

Tab. 6.7: Ergebnisse der Prävalenzschätzung (Bayern) für beide Geschlechter: Prozentangaben vertikal

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
BEI DIAGNOSESTELLUNG	42460	555	612	1596	2151	3211	3564	4619	5814	7899	6770	4609
NACH 0 MONATEN FOLLOW-UP	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NACH 3 MONATEN FOLLOW-UP	97.61	96.48	97.92	99.14	98.11	97.41	97.41	97.41	97.41	97.41	97.41	97.41
NACH 6 MONATEN FOLLOW-UP	91.65	91.89	91.99	95.54	94.75	93.14	92.54	92.54	92.54	92.54	92.54	92.54
NACH 9 MONATEN FOLLOW-UP	87.83	89.11	88.24	92.66	91.62	90.24	89.64	89.64	89.64	89.64	89.64	89.64
NACH 12 MONATEN FOLLOW-UP	76.58	81.17	80.24	88.14	86.14	84.14	83.14	83.14	83.14	83.14	83.14	83.14
NACH 15 MONATEN FOLLOW-UP	71.16	76.17	75.24	82.14	80.14	78.14	77.14	77.14	77.14	77.14	77.14	77.14
NACH 18 MONATEN FOLLOW-UP	64.51	70.11	69.24	76.14	74.14	72.14	71.14	71.14	71.14	71.14	71.14	71.14
NACH 21 MONATEN FOLLOW-UP	58.39	64.11	63.24	70.14	68.14	66.14	65.14	65.14	65.14	65.14	65.14	65.14
NACH 24 MONATEN FOLLOW-UP	54.13	60.11	59.24	66.14	64.14	62.14	61.14	61.14	61.14	61.14	61.14	61.14
NACH 27 MONATEN FOLLOW-UP	49.72	56.10	55.24	62.14	60.14	58.14	57.14	57.14	57.14	57.14	57.14	57.14
NACH 30 MONATEN FOLLOW-UP	45.59	52.11	51.24	58.14	56.14	54.14	53.14	53.14	53.14	53.14	53.14	53.14
NACH 33 MONATEN FOLLOW-UP	41.29	48.10	47.24	54.14	52.14	50.14	49.14	49.14	49.14	49.14	49.14	49.14
NACH 36 MONATEN FOLLOW-UP	37.59	44.11	43.24	50.14	48.14	46.14	45.14	45.14	45.14	45.14	45.14	45.14
NACH 39 MONATEN FOLLOW-UP	33.47	40.11	39.24	46.14	44.14	42.14	41.14	41.14	41.14	41.14	41.14	41.14
NACH 42 MONATEN FOLLOW-UP	27.97	34.11	33.24	42.14	40.14	38.14	37.14	37.14	37.14	37.14	37.14	37.14

Tab. 6.8: Ergebnisse der Prävalenzschätzung (Bayern) für das männliche Geschlecht: Prozentangaben horizontal kumuliert, ersichtlich ist hieraus auch der Beitrag der untersten, nicht dargestellten Altersklassen

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
NACH 0 MONATEN FOLLOW-UP	22068	5,42	4,80	4,48	3,88	3,00	2,20	1,60	1,20	0,80	0,40	0,20
NACH 3 MONATEN FOLLOW-UP	22421	5,27	4,65	4,33	3,73	2,85	2,05	1,45	1,05	0,65	0,35	0,15
NACH 6 MONATEN FOLLOW-UP	18894	4,35	3,73	3,41	2,81	2,11	1,51	1,11	0,71	0,41	0,21	0,11
NACH 9 MONATEN FOLLOW-UP	17494	4,18	3,56	3,24	2,64	1,94	1,34	0,94	0,54	0,24	0,14	0,04
NACH 12 MONATEN FOLLOW-UP	16195	4,12	3,50	3,18	2,58	1,88	1,28	0,88	0,48	0,18	0,08	0,02
NACH 15 MONATEN FOLLOW-UP	14519	4,50	3,88	3,56	2,96	2,26	1,66	1,26	0,86	0,46	0,26	0,12
NACH 18 MONATEN FOLLOW-UP	13273	4,88	4,26	3,94	3,34	2,64	2,04	1,44	1,04	0,64	0,34	0,16
NACH 24 MONATEN FOLLOW-UP	11973	6,89	6,27	5,95	5,35	4,65	3,85	3,05	2,25	1,45	0,85	0,45
NACH 27 MONATEN FOLLOW-UP	10768	6,29	5,67	5,35	4,75	4,05	3,25	2,45	1,65	0,85	0,45	0,25
NACH 30 MONATEN FOLLOW-UP	9781	5,82	5,20	4,88	4,28	3,58	2,78	2,18	1,38	0,78	0,38	0,18
NACH 33 MONATEN FOLLOW-UP	7795	4,93	4,31	3,99	3,39	2,69	1,89	1,29	0,69	0,29	0,14	0,04
NACH 42 MONATEN FOLLOW-UP	7685	5,87	5,25	4,93	4,33	3,63	2,83	2,23	1,43	0,83	0,43	0,23
NACH 45 MONATEN FOLLOW-UP	6512	6,71	6,09	5,77	5,17	4,47	3,67	2,87	2,07	1,27	0,67	0,27
NACH 54 MONATEN FOLLOW-UP	5805	6,82	6,20	5,88	5,28	4,58	3,78	2,98	2,18	1,38	0,78	0,38
NACH 60 MONATEN FOLLOW-UP	4832	6,32	5,70	5,38	4,78	4,08	3,28	2,48	1,68	0,88	0,48	0,28

Tab. 6.9 Ergebnisse der Prävalenzschätzung (Bayern) für das weibliche Geschlecht: Prozentangaben horizontal kumuliert, ersichtlich ist hieraus auch der Beitrag der untersten, nicht dargestellten Altersklassen

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
NACH 0 MONATEN FOLLOW-UP	20392	2,48	4,61	7,50	14,11	21,49	31,21	42,68	56,18	71,75	88,47	100,00
NACH 3 MONATEN FOLLOW-UP	19278	2,21	4,04	6,67	14,11	21,91	31,76	43,37	56,82	73,41	89,66	100,00
NACH 6 MONATEN FOLLOW-UP	18233	2,25	4,76	8,85	14,69	22,31	32,32	44,76	57,67	74,05	89,28	100,00
NACH 9 MONATEN FOLLOW-UP	17249	1,27	4,84	7,04	14,67	22,84	32,87	45,70	58,13	74,69	89,67	100,00
NACH 12 MONATEN FOLLOW-UP	16271	3,32	4,83	6,20	15,74	23,16	33,62	46,53	59,02	75,31	89,95	100,00
NACH 15 MONATEN FOLLOW-UP	15172	3,28	5,00	6,45	16,41	23,77	34,16	46,53	56,95	71,72	87,42	100,00
NACH 18 MONATEN FOLLOW-UP	14125	3,12	5,10	6,64	16,68	24,19	34,30	47,28	60,63	75,67	91,29	100,00
NACH 24 MONATEN FOLLOW-UP	13127	3,18	5,20	6,93	16,40	24,86	35,44	48,17	61,74	77,70	91,29	100,00
NACH 27 MONATEN FOLLOW-UP	12217	3,22	5,29	10,17	16,77	25,47	36,36	49,06	62,67	78,44	91,70	100,00
NACH 30 MONATEN FOLLOW-UP	11378	3,25	5,43	10,47	17,10	26,09	37,12	49,96	63,56	79,21	92,12	100,00
NACH 33 MONATEN FOLLOW-UP	10523	3,53	5,52	10,60	17,54	26,59	37,80	50,85	64,43	79,93	92,47	100,00
NACH 42 MONATEN FOLLOW-UP	9544	4,41	5,65	11,06	18,15	27,43	38,56	51,75	65,72	80,98	93,02	100,00
NACH 45 MONATEN FOLLOW-UP	8633	3,84	5,79	11,44	18,75	28,20	40,05	53,45	65,94	82,02	93,53	100,00
NACH 54 MONATEN FOLLOW-UP	7755	5,55	5,92	11,82	19,32	29,14	41,12	54,75	68,19	82,37	94,07	100,00
NACH 60 MONATEN FOLLOW-UP	7033	3,51	6,07	12,22	19,92	30,02	42,71	56,03	69,88	83,52	94,46	100,00

Tab. 6.10: Ergebnisse der Prävalenzschätzung (Bayern) für beide Geschlechter: Prozentangaben horizontal kumuliert, ersichtlich ist hieraus auch der Beitrag der untersten, nicht dargestellten Altersklassen

	TOTAL	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
NACH 0 MONATEN FOLLOW-UP	42640	3,36	4,80	8,56	13,62	21,19	30,29	41,17	54,86	73,66	89,41	100,00
NACH 3 MONATEN FOLLOW-UP	39899	3,43	4,91	8,76	13,92	21,54	30,76	41,76	55,46	74,00	89,72	100,00
NACH 6 MONATEN FOLLOW-UP	37151	3,51	5,03	8,96	14,22	21,91	31,25	42,35	56,05	74,53	90,32	100,00
NACH 9 MONATEN FOLLOW-UP	34743	3,59	5,15	9,17	14,52	22,11	31,74	42,96	56,65	75,06	90,32	100,00
NACH 12 MONATEN FOLLOW-UP	32515	3,57	5,26	9,38	14,82	22,69	32,22	43,55	57,25	75,58	90,32	100,00
NACH 15 MONATEN FOLLOW-UP	29748	3,78	5,38	9,65	15,66	23,20	32,88	44,36	58,04	76,28	90,99	100,00
NACH 24 MONATEN FOLLOW-UP	25047	3,99	5,75	10,24	16,08	24,75	34,22	45,98	59,63	77,83	91,72	100,00
NACH 27 MONATEN FOLLOW-UP	20985	4,39	5,92	10,53	16,52	24,78	34,69	46,82	60,42	78,38	92,47	100,00
NACH 30 MONATEN FOLLOW-UP	21800	4,21	6,07	10,83	16,93	25,32	35,56	47,62	61,26	78,93	92,47	100,00
NACH 33 MONATEN FOLLOW-UP	17100	4,32	6,24	11,13	17,37	25,87	36,25	48,43	61,99	79,57	92,73	100,00
NACH 42 MONATEN FOLLOW-UP	15112	4,57	6,75	12,07	18,70	26,59	37,24	49,66	63,16	81,51	93,20	100,00
NACH 45 MONATEN FOLLOW-UP	13541	4,82	6,89	12,54	19,36	28,17	39,35	51,11	65,46	82,31	94,08	100,00
NACH 60 MONATEN FOLLOW-UP	11735	4,99	7,24	13,01	20,04	29,21	40,41	52,14	66,80	83,17	94,44	100,00

Ermittlung sind für Männer 897, für Frauen 1675 Neuerkrankungen in der höchsten Altersklasse zu addieren. Damit ergibt sich bei Extrapolation der saarländischen Inzidenzen für Männer eine Inzidenz von 438, für Frauen von 387 Neuerkrankungen pro 100.000 der bayerischen Bevölkerung.

Der Vergleich des altersspezifischen Absterbens (vertikale Prozente) zwischen Männern und Frauen bestätigt die bekannte Tatsache, daß männliche Tumorkrankheiten u.a. wegen des prognostisch ungünstigen und zahlenmäßig dominierenden Bronchialkarzinoms eine geringere mittlere Überlebenswahrscheinlichkeit aufweisen als Frauen vergleichbaren Alters, für die die prognostisch günstigeren Tumoren der Mamma und Gebärmutter im Vordergrund stehen.

Mit zunehmenden Alter verschlechtert sich die tumorbedingte Prognose in beiden Geschlechtern nahezu kontinuierlich, da die prognostisch ungünstigeren Tumoren altersabhängig absolut und relativ zunehmen. Dies hat zur Folge, daß beispielsweise der Anteil des Mammarkarzinoms von über 40% bei den 40 bis 45-jährigen Frauen auf unter 10% in der Altersklasse 80-85 zurückgeht. Damit nimmt auch der geschlechtsbedingte Prognoseunterschied im Alter ab.

Die Tabellen 6.8 bis 6.10 beschreiben dieselben Daten, wobei hier horizontal kumuliert wurde. Es fällt auf, daß das Durchschnittsalter der Patienten im Lauf der Nachsorge nicht zunimmt. Für Männer und Frauen gemeinsam (Tab. 6.10) liegt der Altersmedian zu Beginn der Nachsorge bei etwa 68 Jahren. 54,8% aller Patienten sind bei Diagnosestellung jünger als 70 Jahre. Im Lauf der 60monatigen Nachsorge verschiebt sich der Median in die Klasse der bei Diagnosestellung 60 bis 65 Jahre alten Patienten, also etwa um eine Altersklasse. Berücksichtigt man die Alterung um 5 Jahre im Lauf der Nachsorge, so ergibt sich für den tatsächlichen Altersmedian am Ende des 5jährigen Follow-up derselbe Wert wie zu Beginn.

6.2.4 Folgerungen für die Planung von Nachsorgeprogrammen

Die Spaltensumme in den Tabellen 6.5 bis 6.10 enthält die Anzahl bzw. prozentuale Darstellung der Patienten, die pro Jahr den jeweiligen Zeitpunkt im Verlauf der Nachsorge erreichen. Summierung

dieser Spalten liefert die Anzahl der in einem Jahr maximal durchführbaren Nachsorgeuntersuchungen für Männer, Frauen bzw. insgesamt. Tabelle 6.11 faßt die wesentlichen Ergebnisse zusammen.

Die tatsächliche Inanspruchnahme von Untersuchungen hängt von der Bereitschaft zur Teilnahme und der Rate in Remission befindlicher Patienten ab, denn Nachsorge richtet sich vorwiegend an klinisch tumorfreie Patienten. Da die Bereitschaft zur Teilnahme mit zunehmendem Alter vermutlich abnimmt, ist die Verschiebung des Altersspektrums innerhalb des 5-Jahreszeitraums von Wichtigkeit. Für eine umfassende Simulation der Inanspruchnahme wären allerdings zusätzliche Daten erforderlich.

Tab. 6.11: Zusammenfassung: Anzahl der in Bayern pro Jahr anhand der durchgeführten Modellrechnung theoretisch möglichen Nachsorgeuntersuchungen bei 4-3-3-2-2 Terminen

	davon bei Durchführung			
	männlich	weiblich	insgesamt	jünger als 70 Jahre
Neuerkrankungen	22.068	20.392	42.460	55 %
1. Nachsorge (3 Mon.)	20.421	19.278	39.699	55 %
14. Nachsorge (60 Mon.)	4.832	7.003	11.835	53 %
Summe 1.-14. Nachsorge	166.556	180.584	347.140	-

Als ungefähre Schätzgröße könnte von einer durchschnittlichen Tumorfreiheit von 70% der jeweils Lebenden sowie einer Inanspruchnahme von 70% ausgegangen werden. Unter dieser Voraussetzung würde pro Jahr etwa die Hälfte der theoretisch möglichen Nachsorgeuntersuchungen realisiert, d.h. etwa 175.000 Arzt-Patienten-Kontakte wären im Rahmen der Nachsorge in Bayern zu erwarten.

Die unmittelbaren jährlichen Kosten eines für alle Tumorkranken anwendbaren Nachsorgeprogramms würden damit je nach Umfang der angebotenen Leistungen zwischen 1,7 und 35 Mio. DM betragen, wenn Kosten zwischen 10 DM und 200 DM pro Untersuchung zugrunde-

gelegt werden. Hinzu kämen die Folgekosten für Abklärung verdächtiger Befunde. Bei Erhöhung der Untersuchungsfrequenz oder Verlängerung des Nachsorgezeitraums von 5 auf 10 Jahre nach Diagnosestellung würden sich die Gesamtkosten entsprechend erhöhen. Die prognostizierbare Zahl von Nachsorgekontakten bei engmaschigeren, weitmaschigeren und/oder längerfristigen Planungen kann aus den vorliegenden Tabellen entnommen bzw. hochgerechnet werden.

6.3 Prävalenzschätzung für Metastasen und Rezidive im zeitlichen Verlauf am Beispiel Mammakarzinom

Im folgenden wird versucht, aus den Daten des Tumorregisters München (TRM) Verlaufsaussagen für eine ausgewählte Tumordiagnose zu gewinnen und der Nachsorgestrategie des Tumorzentrums München (TZM) gegenüberzustellen. Ausgewählt wurden die Daten zum Mammakarzinom, da hierzu am TRM die größten Fallzahlen vorliegen. Die Nachsorgestrategie des TZM wurde in (53) veröffentlicht.

6.3.1 Datenlage

6.3.1.1 Mammakarzinome im Tumorregister München

Bis Ende 1983 waren im TRM $n = 4337$ Mammakarzinome der Frau dokumentiert. Hierbei handelt es sich um ein Mischkollektiv aus prospektiv und retrospektiv, prolektiv und retrolektiv, primär und nach Metastasierung erfaßten Fällen. Berücksichtigt wurden operierte unilaterale, lokoregionär begrenzte Fälle ohne vorhergehende andere Tumorerkrankung ($n = 3956$). Zur Schätzung der Rezidivierungs- und Metastasierungswahrscheinlichkeit wurde aus diesem Bestand ein 'prospektives Kollektiv' aus Patienten, die anlässlich der Primärtherapie erfaßt wurden, gebildet, für das zur Vermeidung von Selektionseffekten nur klinikweise annähernd vollständige Jahrgänge herangezogen wurden. Für die Jahrgänge links der fetten Linie in Tab. 6.12 ergibt sich ein stetiger, statistisch signifikanter Trend zum besseren Abschneiden mit rückläufigem Diagnosejahr.

Tab. 6.12: "Definition des prospektiven Kollektivs" von Patientinnen mit Mammakarzinom im Tumorregister München. Berücksichtigt wurden klinikintern annähernd vollständige Jahrgänge

		Jahr					Prosp. Kollektiv
		78	79	80	81	82	Gesamt
Klinik	A	7	46	53	50	67	216
	B	8	12	67	129	136	332
	C	23	25	37	75	81	156
	D	14	22	53	47	42	142
	E	0	0	0	4	19	19
	F	2	1	8	4	3	0
Prosp.Kollektiv gesamt		0	46	173	301	345	865

Für $n = 422$, d.h. 48.8% dieses prospektiven Kollektivs liegen keine Verlaufsinformationen vor. Die verbleibenden 443 Patienten, für die Verlaufsangaben verfügbar sind, wurden im Median 13.5, maximal 48 Monate beobachtet. An 61 Patienten wurden bisher Fernmetastasen im Skelett, Lunge/Pleura, Leber oder Mamma diagnostiziert. Wegen der noch kurzen Beobachtungsdauer und der geringen Häufigkeit von Rezidiven und Metastasen bildet dieses Kollektiv allein keine ausreichende Basis zur Schätzung der organ- und zeitabhängigen Metastasierungswahrscheinlichkeit.

Deshalb wurde zusätzlich ein 'retrospektives Kollektiv', ($n=696$) aus Patienten gebildet, die innerhalb der ersten 5 Jahre nach Diagnosestellung von einer Fernmetastasierung in die obengenannten Organe betroffen waren und deren Primärerkrankung nach 1968, also innerhalb der letzten 15 Jahre, aufgetreten ist. Mit diesem Kollektiv sollte zeitabhängig die Organverteilung von Metastasen, d.h. $p_{t,x}$ (Lokalisation = x | Metastasierung erfolgt z.Zt. t) ermittelt werden.

Problematisch bei der Zusammenstellung dieses Kollektivs ist die fachliche Ausrichtung der Kliniken, da nicht alle Metastasen mit gleicher Wahrscheinlichkeit der gleichen Behandlung und damit der gleichen Klinik zugeführt werden. Zur Vermeidung entsprechender systematischer Fehler wurde jeweils der erste, nicht der aktuelle

Metastasenstatus berücksichtigt. Damit kann - unter der Voraussetzung einer retrospektiv sorgfältigen Erfassung - die klinikbedingte Inhomogenität umgangen werden. Tab. 6.13 zeigt die relativen Häufigkeiten der wichtigsten Metastasenlokalisationen im Vergleich zwischen den Fächergruppen. Die angegebenen χ^2 -Tests (verglichen wurden jeweils spaltenweise k x 2 Felder) sind nicht unabhängig.

Tab. 6.13: Gegenüberstellung von Metastasenlokalisation und Fächergruppe des Datenurhebers im 'retrospektiven Kollektiv' zur Überprüfung etwaiger durch die Zusammensetzung der Kliniken bedingter Selektionseffekte (Prozente horizontal, mehrfache Angaben möglich).

Fach	n	Skelett	Lunge	Leber	Mamma
Operativ	106	45,3%	35,8%	8,5%	29,2%
Innere	167	65,9%	30,5%	14,4%	10,8%
Nachsorge	356	57,0%	34,6%	12,1%	18,5%
Radiologie	67	65,7%	22,4%	16,4%	14,9%
χ^2		p < 0,01	p > 0,2	p > 0,3	p < 0,01

Die beobachteten Abweichungen sind überwiegend auf den Beitrag der operativen Kliniken mit häufigerem Befall der kontralateralen Mamma zurückzuführen. Durch Ausschluß dieses Befalls wird die Inhomogenität beim Skelett beseitigt (p = 0,3).

Die Überrepräsentation der kontralateralen Mamma ist im übrigen vereinbar mit der Vorstellung, daß es sich bei den entsprechenden Tumoren zum Teil um unabhängige Zweitmalignome handelt. In den Originaldaten des Tumorregisters wurde der Befall der kontralateralen Mamma teils als Metastasierung, teils als Zweitmalignomkrankung registriert. Da diesbezüglich das Dokumentationsverhalten der einzelnen Kliniken voneinander abweicht, wurde diese Unterscheidung in der vorliegenden Untersuchung ignoriert. Stattdessen wurden alle Befunde der kontralateralen Mamma als Metastasen ausgewertet, wodurch die Metastasierungsrate überschätzt wird. Andererseits könnte die Vermengung von Metastasierung und Zweitmalignomen auch durch Elimination der kontralateralen Mamma in den vorliegenden Daten nicht beobachtet werden - der Fehler liefe nur in die entgegengesetzte Richtung.

Insgesamt kann das retrospektive Kollektiv als homogen und plausibel angesehen werden. Auch ein Vergleich der Organverteilung zwischen prospektivem und retrospektivem Kollektiv führt für keine Metastasenlokalisation zum Widerspruch, wie Tab. 6.14 zeigt. Aus dem prospektiven Kollektiv wurden in dieser Tabelle nur die bereits metastasierten Fälle (n=61) berücksichtigt, um gleiche Verhältnisse wie im retrospektiven Sample herzustellen.

Tab. 6.14: Metastasenlokalisationen im prospektiven und retrospektiven Kollektiv

	Skelett	Lunge	Leber	Mamma
prospekt. Kollektiv, metastasiert (n=61)	30/61	22/61	5/61	14/61
retrospekt. Kollektiv (n=696)	405/696	227/696	87/696	125/696
χ^2	p > 0,15	p > 0,5	p > 0,3	p > 0,3

Prospektives und retrospektives Kollektiv umfassen zusammen n=1561 Patienten, also ca. 40% der registrierten Fälle, wobei für die Auswertung letztlich nur 1139 oder 29% zur Verfügung standen, da im prospektiven Kollektiv nur jeder 2. Fall Verlaufsdaten aufweist. Diese vergleichsweise geringe Ausbeute ist auf die bisher noch unvollständige Erfassung der Krankheitsverläufe im TRM zurückzuführen.

6.3.1.2 Vergleichskollektiv des Istituto Tumori, Mailand

Ergänzend wurde ein Kollektiv des Mailänder Istituto Nazionale Tumori aus den Jahren 1964-68 (65) vom Umfang n=716 herangezogen. Dieses 'Mailänder Kollektiv' entstand aus der Sekundäranalyse einer randomisierten Studie, in der radikale vs. erweitert radikale Mastektomie geprüft wurden. Es wurde ausgewählt, weil die chirurgische Therapie in beiden Armen nach heutigem Stand als gleichwertig anzusehen ist und keine adjuvante Therapie durchgeführt wurde, die in neueren Studien die dominierende Rolle spielt.

6.3.1.3 Repräsentativität der Daten

Für beide Kollektive wurde zuerst die klinikinterne Selektion überprüft. Für das Mailänder Sample ist mit der für randomisierte Studien typischen Vorselektion, wie z.B. nach Wohnsitz der Patienten, zu rechnen. Gemäß Studienprotokoll wären darüberhinaus nur Patientinnen mit klinisch unauffälligem Lymphknotenstatus in der Axilla zu rekrutieren gewesen. Tatsächlich jedoch wiesen 381/716, also 53% der Patienten, histologisch gesicherte Lymphknotenmetastasen in der Axilla auf. Die Vergleichszahl für das prospektive Kollektiv des Tumoregisters München lautet 351/729, also 48%. Damit sind beide Kollektive im wichtigsten der bekannten prognostischen Kriterien vergleichbar ($p > 0,05$). Andere konkrete Hinweise auf zentrumsinterne Selektion des Mailänder Kollektivs wurden nicht publiziert.

Für das prospektive Kollektiv des Tumoregisters München wurden nur annähernd vollständige Klinikjahrgänge berücksichtigt, um zentrumsinterne Selektion weitgehend zu eliminieren. Zur Abschätzung möglicher Artefakte durch unvollständiges Follow-up wurden zusätzlich die Teilkollektive mit bzw. ohne Verlaufsangabe hinsichtlich der möglicherweise prognoserelevanten Faktoren Stadium, adjuvante Behandlung, Alter, Menopausenstatus, Histologie und Rezeptorstatus verglichen. Alle Kriterien waren unauffällig ($p > 0,2$), so daß gegen die Annahme der Unabhängigkeit von Datenverfügbarkeit und Prognose keine Einwände erkennbar sind.

Zur Frage der Repräsentativität beider Samples kann nur vom Altersaufbau ausgegangen werden. Ausgehend vom Saarländischen Register (60), ergibt sich für die Gesamtbevölkerung die in Tab. 6.15 wiedergegebene Altersverteilung bei Diagnosestellung eines Mammakarzinoms.

Für eine trunkierte, der mutmaßlichen Situation in der Nachsorge entsprechende Bevölkerung bis unter 75 Jahren beträgt der Altersmedian also ca. 57 Jahre. Dies entspricht der Zusammensetzung des prospektiven Kollektivs des TRM, dessen Median bei 55 Jahren liegt, und unterscheidet sich geringfügig vom Mailänder Kollektiv mit 52,5 Jahren.

Tab. 6.15: Altersverteilung bei Diagnosestellung eines Mammakarzi-
noms für die saarländische weibliche Bevölkerung 1981

Altersklasse: ...bis unter	30	35	40	45	50	55	60	65	70	75	75+	Gesamt
absolut	6	12	9	43	54	46	43	56	53	67	106	495
relativ	1.2	2.4	1.8	8.7	10.9	9.3	8.7	11.3	10.7	13.5	21.5	495
kumuliert	1.2	3.6	5.4	14.1	25.0	34.3	43.0	54.3	65.0	78.5	100.0	495
relativ, trunziert	1.5	3.1	2.3	11.1	13.9	11.8	11.1	14.4	13.6	17.2	-	389
kumuliert, trunziert	1.5	4.6	6.9	18.0	31.9	43.7	54.8	69.2	82.8	100.0	-	389

Über die Altersverteilung hinaus bestehen keine Möglichkeiten zur direkten Beurteilung der Repräsentativität, da für die Gesamtbevölkerung keine differenzierteren Daten vorliegen.

Dies gilt auch für den axillären Lymphknotenbefall, dessen prognostische Bedeutung in beiden Kollektiven reproduziert wurde.

6.3.2 Organ- und zeitabhängige Metastasierungswahrscheinlichkeiten

6.3.2.1 Methodik der Schätzung

Die verstärkte Nutzung verlaufsorientierter Tumorregister läßt auf lange Sicht eine Erweiterung des Wissens zum Krankheitsverlauf erwarten. Allerdings zeigt die Strukturanalyse des auf den ersten Blick umfangreichen Materials im TRM, daß die verlaufsorientierten Tumorregister in der Bundesrepublik - denn diese Beobachtung gilt nicht nur für München - noch mindestens 3-4 Jahre benötigen, bis die Ableitung von Verlaufsaussagen allein aus der prospektiven Dokumentation möglich wird.

In der jetzigen Situation wurde versucht, Verlaufsaussagen durch Verbindung von prospektivem und retrospektivem Kollektiv und Abgleich mit publizierten Daten zu gewinnen. Hierbei wurde zweistufig vorgegangen. Im ersten Schritt wird aus den Daten des prospektiven Kollektivs des TRM sowie des Mailänder Kollektivs das zeitliche Auftreten von Organmetastasen insgesamt geschätzt. Im zweiten Schritt wird diese globale Metastasierungswahrscheinlichkeit mit den verschiedenen, aus dem retrospektiven Kollektiv des TRM gewichteten Zielorganen verknüpft, um damit organspezifische

Metastasierungskurven zu erhalten. Durch diese Zweiteilung konnte insbesondere die breite empirische Basis des TRM über die relativen Häufigkeiten einzelner Metastasenlokalisationen bestmöglich genutzt werden.

Schätzung des zeitlichen Auftretens von Metastasen insgesamt:

verwendet wurden die Daten des prospektiven Kollektivs des TRM sowie des Mailänder Kollektivs. Unter der Annahme exponentialverteilter Remissionsdauern, d.h. $S(t) = \exp(-\lambda t)$, wurden Schätzer für λ berechnet. Für das Mailänder Kollektiv waren hierbei nur Drei- bzw. Fünfjahresraten verfügbar, so daß die Schätzung von λ über die Modellgleichung x-Jahres-Rate = $(-\lambda_1 \times \text{Jahre})$ erfolgen mußte. Für das prospektive Kollektiv des Tumorregisters lagen Einzelfalldaten vor, für die der Maximum-Likelihood-Ansatz

$$\lambda_2 = d / \sum_{i=1}^N t_i$$

d = Anzahl beobachteter Zielereignisse
 t_i = Zeit bis zum Ende der Beobachtung

anwendbar ist.

Aus diesen über unterschiedliche Ansätze gewonnenen Schätzern λ_i wurde ein Schätzer λ gemittelt. Dieser dient als Basis der anschließenden Berechnungen.

Ermittlung der organspezifischen Metastasierungsraten:

Im Hinblick auf das Ziel von Nachsorgeprogrammen, die Überwachung klinisch tumorfreier Patienten, wurde nur der erste Metastasenstatus am Ende der Remission berücksichtigt. Denn nach manifester Metastasierung kehrt der Patient nicht mehr in die reguläre Nachsorge zurück, so daß später hinzukommende weitere Metastasen für die vorsorgende Betreuung in Nachsorgeprogrammen bedeutungslos sind. Dagegen führt das Auftreten eines lokoregionären Rezidivs ohne gleichzeitige Fernmetastasierung nicht zum Ende der Beobachtung, da für Rezidive eine potentiell kurative Behandlung existiert und Rückkehr in die Routinenachsorge nicht auszuschließen ist.

Zur Erstellung der organspezifischen Metastasierungskurven für Skelett-, Lungen-, Leber- und Mammametastasen wurde aus dem retrospektiven Kollektiv für jedes Zeitintervall nach Primärtherapie der Anteil der jeweiligen Metastasenlokalisation am Gesamtaufkommen an Metastasen geschätzt, also die bedingte Wahrscheinlichkeit

$$p_{x,t} = p(\text{Metastasierung im Organ } x | \text{Metastasierung erfolgt im Zeitintervall } t)$$

Die Auftretenswahrscheinlichkeit $m_{x,t}$ der Metastase x bis zum Zeitpunkt t ergibt sich damit durch

$$m_{x,t} = (1-S(t)) \cdot p_{x,t}.$$

Bei Zeitunabhängigkeit von $p_{x,t}$, d.h. $p_{x,t} = p_x$,

vereinfacht sich die Berechnung zu

$$m_x = (1-S(t)) \cdot p_x.$$

Auftreten lokoregionärer Rezidive:

Das zeitliche Auftreten von lokoregionären Rezidiven könnte mit der gleichen Methode geschätzt werden. Allerdings ist hierbei zu berücksichtigen, daß die Häufigkeit lokoregionärer Rezidive in der Literatur stark variiert. Vor einer quantitativen Schätzung von Rezidivraten ist zu prüfen, ob diese aufgrund der weltweit publizierten Daten plausibel durchführbar ist.

6.3.2.2 Schätzung der metastasenfreien Zeit

Die für die Erstellung einer lokalisationsunabhängigen Modellkurve für das zeitliche Auftreten von Fernmetastasen nach potentiell kurativer Primärtherapie verfügbaren Daten sind in Abbildung 6.3 dargestellt.

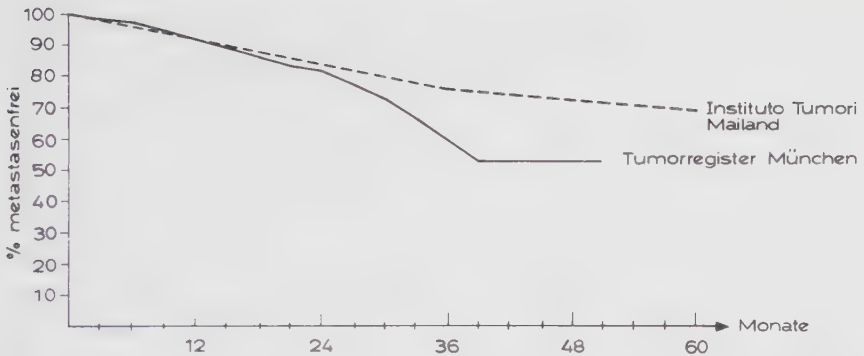


Abb. 6.3: Beobachteter Verlauf der Fernmetastasierung für das prospektive Kollektiv des Tumorregisters München (n=445) und das Mailänder Kollektiv (n=716)

Im TRM ergaben sich überwiegend kurze Beobachtungszeiten. Bereits nach 12 Monaten waren 40% der Patienten nicht mehr im Follow-up, nach 24 Monaten standen nur noch 23%, nach 3 Jahren noch 4% unter Risiko. Nach 4 Jahren schließlich lagen nur noch für eine Patientin Verlaufsdaten vor.

Tab. 6.16: Schätzung für λ (Zeiteinheit=Tage) zur Beschreibung der metastasenf freien Zeit bzgl. Skelett, Lunge/Pleura, Leber, kontralateraler Mamma

Quelle	Zeitraum		geschätzte 5-Jahresrate
Mailand	3 Jahre	0.000251	63%
Mailand	5 Jahre	0.000203	69%
Tumorregister	3 Jahre	0.000278	60%
Tumorregister	4 Jahre	0.000285	59%
(Tumorregister	5 Jahre	0.000285	59%)
Festlegung	-	0.00025	63%

Die aus den Daten des TRM abgeleiteten Modellkurven unterscheiden sich in der 5-Jahresrate um ca. 10%-Punkte von der geschätzten (= beobachteten) 5-Jahresrate des Mailänder Kollektivs. Hierbei muß berücksichtigt werden, daß die Mailänder Daten als Ergebnis einer randomisierten Studie eher eine Überschätzung als eine Unterschätzung der tatsächlichen Ereignisfreiheit erwarten lassen.

Demgegenüber sind im TRM selektionsbedingte Effekte aufgrund des unvollständigen Follow-up nicht auszuschließen bzw. wahrscheinlich. Es ist keineswegs gewährleistet, daß die beobachtete Zensierung als Zufallsmechanismus gedeutet werden kann. Insbesondere kann nicht ausgeschlossen werden, daß die Wiedervorstellung von Patientinnen vielfach nach Einsetzen der Progression durch gezielte Zuweisung von außen erfolgt ist. Falls dieser Zensierungsmechanismus am Zustandekommen der Daten beteiligt war, liegt der Erwartungswert des tumorfreien Intervalls für die aus dem Follow-up verlorenen Patientinnen höher als für die unter Beobachtung der Kliniken stehenden Patientinnen. Auch dies gehört zu den Anlaufschwierigkeiten eines versorgungsbezogenen Tumorregisters.

Die Festlegung $\lambda = 0.00025$, entsprechend einer 5-Jahresrate der Metastasenfreiheit von 63% als Grundlage der nachfolgenden Modellrechnungen scheint beiden Fehlerquellen gerecht zu werden.

6.3.2.3 Schätzung des Auftretens der verschiedenen Organmetastasen im zeitlichen Verlauf

Anhand der beobachteten relativen Häufigkeiten der einzelnen Organmetastasen aus dem retrospektiven Kollektiv und der geschätzten globalen Hazard-Funktion $\lambda(t) = \lambda = 0,00025$ wurden zeitabhängige Metastasierungswahrscheinlichkeiten für die einzelnen Metastasenlokalisationen entwickelt.

Den Ausgangspunkt bildet die Beschreibung der bedingten Wahrscheinlichkeiten $p(M_x | \text{Metastasierung erfolgt zur Zeit } t)$ des Auftretens der Metastase x zur Zeit t , unter der Voraussetzung, daß zur Zeit t irgendeine Fernmetastasierung auftritt. M_x bezeichnet das Auftreten der Metastase x , gegebenenfalls gemeinsam mit weiteren Metastasen.

Tab. 6.17 zeigt den zeitlichen Verlauf der relativen und absoluten Häufigkeiten von Skelett-, Lunge-/Pleura-, Leber- und Mammametastasen. Die Abfolge der relativen Häufigkeiten zeigt für keine der Lokalisationen einen steigenden oder fallenden Trend. Die z -Werte nach Wallis/Moore bewegen sich zwischen 0.1 und 1.78, zu beurteilen anhand der Standardnormalverteilung. Auch die Korrelation der relativen Häufigkeiten mit der Zeit erbringt unauffällige Koeffizienten.

Tab. 6.17: Verlauf der relativen Häufigkeit der einzelnen Metastasenlokalisationen im 5-Jahreszeitraum nach Primärtherapie, geschätzt im retrospektiven Kollektiv. Mehrfachangaben bei synchronem Auftreten mehrerer Metastasen

Zeit (Tage)	N	N	p	N	p	N	p	N	p
		Skelett	Skelett	Lunge	Lunge	Leber	Leber	Mamma	Mamma
- 91	23	9	.39	4	.17	3	.13	11	.48
- 182	40	28	.70	9	.23	4	.10	6	.15
- 273	52	24	.46	20	.38	8	.15	13	.25
- 365	54	28	.52	21	.39	4	.07	11	.20
- 456	54	30	.56	16	.30	9	.17	9	.17
- 547	48	31	.65	13	.27	5	.10	6	.13
- 638	43	31	.72	15	.35	5	.12	7	.16
- 730	54	31	.57	21	.39	8	.15	7	.13
- 821	42	23	.55	18	.43	6	.14	5	.12
- 912	38	26	.68	6	.16	5	.13	6	.16
- 1003	40	24	.60	12	.30	6	.15	7	.18
- 1095	40	23	.58	12	.30	6	.15	10	.25
- 1277	57	36	.63	22	.39	5	.09	5	.09
- 1460	40	26	.65	13	.33	6	.15	6	.15
- 1642	40	18	.45	11	.28	4	.10	11	.28
- 1825	31	17	.55	14	.45	3	.10	5	.16
Summe	696	405		227		87		125	
r (Zeit)	-	-	0.09	-	0.33	-	-0.09	-	-0.28
z (Wallis & Moore)	-	-	1.15	-	1.78	-	0.1	-	0.52

Im folgenden wird deshalb von der Konstanz der relativen Häufigkeit im Verlauf der Nachsorge ausgegangen. Der Anteil der Metastasenlokalisation x am Gesamtaufkommen wird folgerichtig als zeitunabhängige Größe $P_x = N_x/N$ geschätzt. Vergleichswerte für das Mailänder Kollektiv sind aufgrund der in (65) gewählten Darstellung nicht ableitbar. Auffällig ist der möglicherweise überhöhte Anteil des Befalls der kontralateralen Mamma, dessen Interpretation und weitere Behandlung bereits weiter oben diskutiert wurden.

Da sich keine Einwände gegen die Annahme konstanter Metastasierungsmuster über die Zeit ergaben, erfolgt die Schätzung der organspezifischen Metastasierungsraten auf der Basis einer proportionalen Gewichtung der globalen Hazard-Rate $\lambda = 0.00025$.

Tab. 6.18: Zeitunabhängige relative Häufigkeiten der einzelnen Metastasenlokalisationen x

Metastasenlokalisation _x	N _x	N _x /N
Skelett	405	.58
Lunge	227	.33
Leber	87	.13
Mamma	125	.18

Da zur frühzeitigen Entdeckung einer Metastase je nach ihrer Lokalisation unterschiedlich aufwendige Maßnahmen erforderlich sind, ist auch das Muster des Miteinanderauftretens verschiedener Metastasen von Interesse. Tab. 6.19 beschreibt die Wahrscheinlichkeit für das Vorliegen weiterer Befunde unter der Bedingung des Vorliegens einer der drei selteneren Metastasen.

Tab. 6.19: Muster des gemeinsamen (synchronen) Auftretens von Metastasen: $p(\text{Lok B}|\text{Lok A})$

Lokalisation B						
Lokalisation A	Skelett	Lunge	Leber	Mamma	keine	n
Skelett	-	0.19	0.12	0.01	0.73	405
Lunge	0.34	-	0.15	0.02	0.59	227
Leber	0.56	0.38	-	0.00	0.26	87
Mamma	0.05	0.03	0.00	-	0.94	125

Die Wahrscheinlichkeit $p(\text{Skelett}/\text{Leber})$ wird als 0.56 abgelesen, d.h. an 56% der Patientinnen, die Lebermetastasen entwickelt haben, werden gleichzeitig Skelettmetastasen diagnostiziert. Nur jede vierte Patientin (26%), die überhaupt Lebermetastasen bereits zu Beginn der Metastasierung aufweist, erkrankt ausschließlich an Lebermetastasen.

Die nahezu verschwindende Abhängigkeit zwischen dem Befall der kontralateralen Mamma und anderweitiger Organmetastasierung deutet erneut darauf hin, daß in der Mehrzahl der Fälle Zweitmalig-nome anzunehmen sind.

6.3.2.4 Lokoregionäre Rezidive

Die Nachbehandlung des radikal operierten Mammakarzinoms ist un-einheitlich, sowohl hinsichtlich der Indikationsstellung zur Nachbehandlung an sich, als auch bezüglich der Wahl der Behand-lungsart nach erfolgter Indikationsstellung zur Nachbehandlung. Zur Wahl stehen Nachbestrahlung und adjuvante Chemo- bzw. Hormon-therapie. Wirkungsweise und Zielvorstellungen für lokale vs. sy-temische Nachbehandlung sind unterschiedlich. Ziel der adjuvan-ten systemischen Therapie ist die frühzeitige Elimination von Mi-krometastasen im ganzen Körper. Demgegenüber zielt die Nachbe-strahlung ausschließlich auf die Sanierung des Entstehungsgebiets des Mammakarzinoms ab. Im Gegensatz zur adjuvanten Chemotherapie, die erst seit einigen Jahren diskutiert wird, geht die Tradition der Nachbestrahlung bis zum Jahr 1901 zurück, wo erstmals eine Patientin nachbestrahlt wurde (36).

Trotz dieser jahrzehntelangen Erfahrung mit der Nachbestrahlung ist die Indikationsstellung zur Nachbestrahlung vs. alleiniger Operation nicht standardisiert. Beispielsweise stimmen die Mit-glieder der Projektgruppe Mammakarzinom im TZM lediglich in der prognostisch günstigsten Untergruppe (pT1 pN0 M0, Rezeptor - positiv) darin überein, auf Nachbehandlung zu verzichten. Anson-sten bestimmt für nodal negative Patientinnen der Stadien pT1 - pT3 die Wahl der Klinik, ob nachbehandelt wird oder nicht.

Ist die Entscheidung zur Nachbehandlung gefallen, so werden im Raum München derzeit alternativ verschiedene Behandlungsstrate-gien praktiziert bzw. empfohlen. Die folgende Abbildung be-schreibt die Situation, wobei

0 = keine Nachbehandlung

⚡ = Bestrahlung

TAM = Tamoxifen

CMF = adjuvante Chemotherapie nach BONADONNA

AC = adjuvante Chemotherapie nach SALMON/JONES

bezeichnen.

POSTOPERATIVE BEHANDLUNG DES MAMMAKARZINOMS DER FRAU NACH EINGESCHRÄNKTER RADIKALER OPERATION

Krankheitsstadium	Rezeptorstatus	Menopausenstatus	häufig durchgeführt	BEHANDLUNGSMÖGLICHKEITEN	
				Mögliche Alternative	in Erprobung (z.B. Studien)
pT1 pNo Mo lateral	-	prä und post	Ø		21 cm: 6xCMF
pT1 pNo Mo med./zentr.	-	prä und post	Ø	↓	21 cm: 6xCMF
pT2 pNo Mo lateral	+ und -	prä post post	Ø Ø Ø	↓ ↓ ↓	6xCMF 6x CMF TAM oder 6xCMF+TAM
pT2 pNo Mo med./zentr.	+ und -	prä post post	Ø oder ↓ Ø oder ↓ Ø oder ↓	↓ ↓ ↓	6xCMF 6xCMF TAM oder 6xCMF+TAM
pT3 pNo Mo	+ und -	prä Post post	↓ oder Ø ↓ oder Ø ↓ oder Ø		6xCMF 6xCMF TAM oder 6xCMF+TAM } (2 ↓)
pT4 pNo Mo	+ und -	prä Post post	↓ ↓ ↓		6xCMF 6xCMF TAM oder 6xCMF+TAM } (2 ↓)
pT1-3a pN1 (≤ 3 Lkn) Mo	- + - +	prä prä post post	6xCMF 6xCMF ↓ ↓	↓ ↓ 6xCMF 6xCMF+TAM	6xCMF+↓ TAM+↓ oder 6xCMF+↓ 6xCMF+↓ TAM+↓ oder 6xCMF+TAM+↓
pT1-3a pN1-2 (≥ 3 Lkn) Mo	+ und -	prä post post	↓ ↓ ↓	6xAC oder 6xCMF 6xAC oder 6xCMF 6xAC+TAM oder 6xCMF+TAM	6xAC+↓ oder 6xCMF+↓ 6xAC+↓ oder 6xCMF+↓ 6xAC+TAM+↓ oder 6xCMF+TAM+↓
Häufig nicht radikal operable Krankheitsstadien					
T1-3a N3 Mo T3b-4 N1-3 Mo	+ und -	prä und post	↓		↓ +6xCMF+TAM ↓ +6xAC+TAM

Abb. 6.4: Postoperative Behandlung des Mammakarzinoms der Frau nach eingeschränkter radikaler Mastektomie im Tumorzentrum München (53)

Der Einfluß der Nachbestrahlung auf die Häufigkeit lokoregionärer Rezidive kann als gesichert gelten. Die wichtigsten Studien sind in (36) zusammengestellt. Allerdings führt die Senkung des Rezidivrisikos anscheinend nicht zu einer Verbesserung der Überlebenswahrscheinlichkeit. Einen Einfluß der Nachbestrahlung auf die Überlebensrate konnte von keiner der großen Studien nachgewiesen werden. Die bloße Senkung des Rezidivrisikos wird nicht von allen Kliniken als hinreichende Rechtfertigung zur Durchführung einer Nachbestrahlung angesehen. Daher wird die Indikationsstellung zur Nachbestrahlung uneinheitlich gehandhabt.

Für die Praxis folgt, daß derzeit mit lokoregionären Rezidiven in unterschiedlicher Frequenz gerechnet werden muß. Je nach praktizierter Nachbehandlung sind Rezidivquoten von 2,5 bis 25% innerhalb von 5 Jahren zu erwarten (30). Die Frequenz des Zielereignisses lokoregionäres Rezidiv hängt somit überwiegend von der Art der Nachbehandlung ab, die derzeit angesichts der Zusammenstellung in Abbildung 6.4 für den Einzugsbereich eines Nachsorgeprogramms nicht vorhersehbar ist. Deshalb wurde auf die Schätzung des zeitlichen Auftretens lokoregionärer Rezidive verzichtet.

Soweit schließlich die frühzeitige Entdeckung von Fernmetastasen unterstützt werden soll, ist der direkte Einfluß schwer meßbar. Indirekte Einflüsse der klinischen Beobachtung eines Patienten auf die Ergebnisse der apparativen Diagnostik bestehen. Im Abschnitt 3.3.2.3 wurden Möglichkeiten geschildert, wie die diagnostische Ausbeute der eingesetzten Tests durch die Organisation eines Nachsorgeprogramms gefördert werden kann.

Ein spezieller Beitrag regelmäßig durchgeführter klinischer Untersuchungen zur Früherkennung einer Progression ist die Entdeckung klinisch nicht mehr stummer Metastasen frühzeitig nach Einsetzen der Symptome. Insofern schafft evtl. bereits die klinische Untersuchung von Patienten allein günstigere Verhältnisse als bei völligem Verzicht auf die Nachsorge. Bei Vorliegen von Symptomen wird in (53) die außerplanmäßige Durchführung von Tests empfohlen.

Die apparativen Untersuchungen werden wohl nur teilweise zur Früherkennung asymptomatischer Tumorprogression eingesetzt, während zum Teil bereits durch Anamnese bzw. klinische Untersuchung ein Verdacht entstanden ist, der durch die apparative Diagnostik abgeklärt wird. Ein Teil der Ausbeute eines Nachsorgeprogramms geht somit nur scheinbar auf die apparative Komponente des Programms zurück, während in Wirklichkeit die klinische Untersuchung allein ausgereicht hat, den Verdacht zu äußern. Dieser Effekt kann im folgenden nicht berücksichtigt werden. Stattdessen werden alle Patienten, die sich einem diagnostischen Test unterziehen, als asymptomatisch und vor Testdurchführung frei von Metastasenverdacht angesehen. Dadurch wird der Beitrag des Tests potentiell überschätzt.

6.3.3.2 Methoden und Annahmen zur Gegenüberstellung

In Abschnitt 6.3.2 wurden Modellkurven entwickelt, durch die das zeitliche Auftreten beobachteter Organmetastasen beschrieben wird. Vor der Anwendung dieser Modellkurven auf eine Nachsorgestrategie ist jedoch die Präzisierung der durch diese Kurve beschriebenen Zielereignisse erforderlich. Denn aus den Daten ist zwar der Zeitpunkt ersichtlich, von dem an der behandelnde Arzt von der Existenz eines Tumorbefalls in dem betreffenden Organ ausging. Zu diesem Zeitpunkt ist das Ereignis Metastasierung registriert. Es ist jedoch nicht erkennbar, auf welcher Information dieser Befund fußt.

Ingesamt beschreiben die beobachteten Daten die Realität nur unzureichend. Falsch-negative und falsch-positive Befunde sind nicht auszuschließen, und im Falle eines richtig-positiven Befundes ist unbekannt, in welchem Entwicklungsstadium sich die betreffende Metastase zum Zeitpunkt ihrer Entdeckung befand. Klinisch manifeste Befunde mit massiver Symptomatik und latente Metastasenherde mischen sich in unbekanntem Verhältnis. Die beobachtete Zeitverteilung ist das gewogene Mittel der Zeitverteilungen für klinische Manifestation und frühzeitige Diagnosestellung mit unbekanntem Anteil beider Gruppen. Wie groß die resultierende zeitliche Unschärfe ist, wird durch einen weiteren Faktor bestimmt, nämlich die Verteilung der dem Metastasierungsmuster zugrundeliegenden Wachstumsraten. Auch hierüber geben die verfügbaren Daten keinen unmittelbaren Aufschluß. In dieser Situation wurde eine einfache, der Datenlage angepaßte Approximation gewählt.

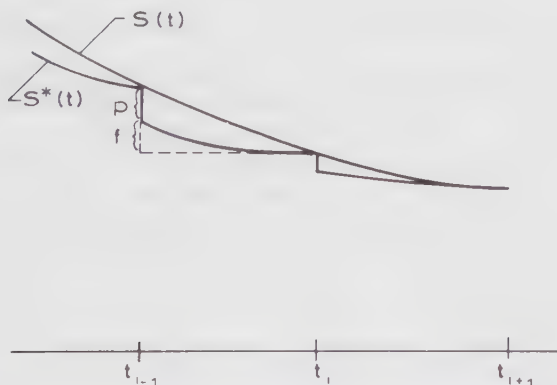


Abb. 6.6: Vereinfachter Ansatz zur a-priori-Evaluation einer Nachsorgestrategie. Dargestellt sind die aus den empirischen Daten ermittelte Modellkurve $S(t)$ und die durch apparative Nachsorge entstehende Kurve $S^*(t)$. Die t_i bezeichnen die relevanten Zeitpunkte des Nachsorgeprogramms, an denen der Test eingesetzt wird.

Es wird angenommen, wie in Abb. 6.6 skizziert, daß jeder der zum Zeitpunkt t_i durch die Nachsorge erfaßten Befunde unter den bisherigen Verhältnissen, aus denen $S(t)$ geschätzt wurde, im Intervall $[t_i, t_{i+1})$ aufgetreten wäre. Für jeden Zeitpunkt t_i ergibt sich $S(t_i) = S^*(t_i)$, d.h. S und S^* kommen jeweils unmittelbar vor

der Ausführung eines Nachsorgeprogrammschrittes zur Deckung. Der Anteil $p/(p+f)$ wird als Sensitivität bezeichnet. Auch dies ist simplifizierend, da zum Zeitpunkt t_i nur ein unbekannter Anteil der Befunde $(p+f)$ für den Test erkennbar ist, während gleichzeitig einige der im Intervall $[t_{i+1}, \infty)$ diagnostizierten Metastasen bereits richtig befundet werden könnten.

Diese Annahmen ermöglichen es, das Nachsorgeprogramm in unabhängige Segmente zu zerlegen, die einzeln analysiert bzw. simuliert werden können. Mit den Kurzbezeichnungen

U1 = Röntgen, Szinti, Sono, Mammographie

U2 = Röntgen, Szinti, Sono

U3 = Röntgen, Szinti

U4 = Röntgen

für die im Nachsorgeprogramm definierten Untersuchungsschritte und A1 = Röntgen, Szinti, Mammographie

als nicht im Programm formulierte Alternativstrategie wurden die in Tabelle 6.20 beschriebenen Strategiesegmente simuliert.

Tab. 6.20: Strategiesegmente zur Nachsorge beim Mammakarzinom. Berücksichtigt ist nur apparative Diagnostik

Jahr	Urheber	Nachsorgestrategie		
		Jahres- anfang	Jahres- mitte	Kurzbezeichnung
1-3	Tumorzentrum	U1	U2	P 1-3
1-3	Tumorzentrum	U1	U3	P'1-3
1-3	Hypothetische Alternative	A1	U3	Q 1-3
4-5	Tumorzentrum	U1	U4	P 4-5
4-5	Hypothetische Alternative	A1	U4	Q 4-5

Bei den alternativen Programmsegmenten Q 1-3 und Q 4-5, in denen die Routineuntersuchung der Leber fallen gelassen ist, wird angenommen, daß die Untersuchung beim Auftreten von Beschwerden bzw. anderen Krankheitsmanifestationen nachgeholt wird. Dies wird auch für andere Organe, die nur selten als erster Metastasierungsort in Erscheinung treten, z.B. das ZNS, so praktiziert. Für solche Untersuchungen werden häufig die diagnostisch aussagekräftigsten Verfahren, z.B. CT statt Sonographie, benutzt.

Die Wahrscheinlichkeiten des paarweisen Auftretens von Metastasen, insbesondere auch die Prävalenz an Lebermetastasen bei positivem Skelett- bzw. Lungenbefund, werden aus Tab. 6.19 geschätzt.

6.3.3.3 Ergebnisse der Gegenüberstellung

Unter den erläuterten Voraussetzungen und Annahmen kann die Wirkung der zu prüfenden Nachsorgestrategie allein durch Angabe von Wahrscheinlichkeiten erschöpfend charakterisiert werden. Zur Erhöhung der Anschaulichkeit wird jedoch zusätzlich die bayerische Bevölkerung als Modellpopulation unterlegt. Sämtliche erwarteten Wahrscheinlichkeiten werden simultan als Befundhäufigkeiten dargestellt, wie sie sich bei Anwendung der Nachsorgestrategie des TZM in Bayern und 100%iger Teilnahme der im 5-Jahreszeitraum in Beobachtung stehenden Patientinnen ergeben würden.

Bei einer anhand anderer Register anzunehmenden Inzidenz von ca. 80 Neuerkrankungen pro 100.000 der weiblichen Bevölkerung treten in Bayern bei ca. 5,7 Millionen weiblichen Einwohnern 4570 Erkrankungsfälle pro Jahr auf. Zum Vergleich ergeben sich pro Jahr in Bayern ca. 2300 Sterbefälle an bösartigen Neubildungen der weiblichen Brust (2). Dies entspricht einer Rate von 40 auf 100.000. Andererseits beträgt die Wahrscheinlichkeit, innerhalb von 30 Jahren dem Grundleiden zu erliegen, ca. 57% (56). Auch aus dieser Sicht erscheint die Annahme einer Inzidenz von 80 pro 100.000 realistisch.

Unter der Steady-State-Annahme bezüglich Inzidenz und Bevölkerung entsprechen die im 5-Jahresverlauf einer Jahrgangskohorte beobachteten Befundhäufigkeiten zahlenmäßig den durch Querschnittsbeobachtungen von 5 sequentiellen Jahrgangskohorten über 1 Jahr zu erwartenden Werten. Die in 5 Nachsorgejahren an einer Kohorte auftretenden Befunde beschreiben daher, quasi übereinandergeschoben, den Querschnitt eines Jahres. Tabelle 6.21 gibt die Anzahl der pro Jahr erwarteten Untersuchungen und Ereignisse wieder. Nachsorgetermine, an denen lediglich klinische Untersuchung und Labor vorgesehen sind, wurden nicht berücksichtigt. Die Strategien P1-3, P'1-3 und P3-4 sind in Tab. 6.20 erklärt.

Innerhalb eines Jahres fallen für 19197 Patientinnen 37540 Nachsorgetermine mit apparativer Diagnostik an. Zu jedem Termin erfolgt die Röntgenuntersuchung des Thorax, 31188 Skelettszintigraphien werden angefertigt, 19197 Mammographien und 19197 bis 31188 Oberbauchsonogramme sind jährlich durchzuführen. Durch diese Untersuchungen werden maximal 970 Skelettmetastasen, 552 Lungenmetastasen, 302 Mammametastasen und 216 Lebermetastasen diagnostiziert.

Tab. 6.21: Anzahl der Untersuchungen/Jahr bei Anwendung der vom Tumorzentrum München vorgeschlagenen Nachsorgestrategie und 100%iger Teilnahme in Bayern (Bevölkerung: 5.712.943 weibliche Einwohner)
Angenommene Inzidenz: 80/100.000
Metastasenfreiheit nach exp. (-.00025 x Tage)
Angabe in Klammern: Befundhäufigkeiten im Intervall $[t_i, t_{i+1})$.

"*": Dieses Organ wird zum angegebenen Zeitpunkt untersucht.

Zeitpunkt	n	Skelett	Lunge /Pleura	Mamma	Leber, Strategie P	Leber, Strategie P'
0 Mon.	4570	* (118)	* (67)	* (72)	* (26)	* (52)
6 Mon.	4367	* (113)	* (64)		* (25)	
12 Mon.	4171	* (108)	* (61)	* (65)	* (24)	* (47)
18 Mon.	3986	* (103)	* (59)		* (23)	
24 Mon.	3808	* (98)	* (56)	* (60)	* (22)	* (43)
30 Mon.	3638	* (94)	* (53)		* (21)	
36 Mon.	3476	* (176)	* (51)	* (55)	* (39)	* (39)
42 Mon.	3321		* (49)			
48 Mon.	3172	* (160)	* (47)	* (50)	* (36)	* (36)
54 Mon.	3031		* (45)			
Summe Unters.						
P1-3	24540	24540 (634)	24540 (360)	12549 (197)	24540 (141)	
P'1-3						12549 (142)
P3-4	13000	6648 (336)	13000 (192)	6648 (105)	6648 (75)	6648 (75)
Gesamt	37540	31188 (970)	37540 (552)	19197 (302)	31188 (216)	19197 (217)

Die folgenden Tabellen zeigen die Ergebnisse der 4 betrachteten Tests unter verschiedenen Annahmen zu Sensitivität und Spezifität. Wegen der Engmaschigkeit der angesetzten Termine ist die Prävalenz pro Untersuchung für alle Tests gering. Deshalb ist die Spezifität der Tests die Schlüsselgröße zur Bestimmung der prädiktiven Werte. Eine präzise Festlegung von Sensitivität und Spezifität allein aufgrund publizierter Daten - z.B. (34, 62) zur Präzision der Skelettszintigraphie, (54) zur Treffsicherheit der

Tab. 6.22: Ergebnisse der Gegenüberstellung von Metastasierungsmuster und Nachsorgestrategie: Skelett
Dargestellt sind der prädiktive Wert des positiven Tests und die erwarteten Befundhäufigkeiten in der bayerischen Bevölkerung

Programm	Prävalenz pro Test	Sens.	Spez.	Präd.Wert d.pos.Tests	Befunde pro Jahr in Bayern		
					Gesamt richtig pos. falsch pos.		
P1-3	2.6%	50	50	0.026	634	317	11875
			70	0.043			7055
			90	0.118			2369
		70	50	0.036	634	444	11889
			70	0.059			7081
			90	0.157			2384
		90	50	0.046	634	571	11842
			70	0.074			7145
			90	0.194			2372
		50	50	0.051	336	168	3126
			70	0.082			1880
			90	0.212			624
P4-5	5.1%	70	50	0.070	336	235	3075
			70	0.111			1882
			90	0.273			625
		90	50	0.088	336	302	3129
			70	0.139			1871
			90	0.326			624

Oberbauchsonographie - ist kaum begründbar. Hinzu kommt, daß das Umfeld eines Nachsorgeprogramms andere Vorbedingungen für die Tests schafft, so daß die Übertragbarkeit publizierter Schätzer für Sensitivität und Spezifität zweifelhaft wäre, auch wenn diese von weiteren Autoren bestätigt wären. Dies wurde in Abschn. 3.3.2 diskutiert. Daher werden die erwarteten Resultate unter dem Einfluß unterschiedlicher Eingangsbedingungen berechnet.

Die ungünstigste Annahme einer Spezifität von nur 50% entspricht dem durch die oben zitierten Arbeiten vermittelten pessimistischen Bild. Auch die gegenteilige Annahme einer hohen Spezifität von 90% wäre begründbar. Welche Werte tatsächlich erreicht werden, dürfte nicht zuletzt von der Organisation der Nachsorge bestimmt werden.

Tab. 6.23: Ergebnisse der Gegenüberstellung von Metastasierungsmuster und Nachsorgestrategie: Lunge/Pleura
Dargestellt sind der prädiktive Wert des positiven Tests und die erwarteten Befundhäufigkeiten in der bayerischen Bevölkerung

Programm	Prävalenz pro Test	Sens.	Spez.	Präd.Wert d.pos.Tests	Befunde pro Jahr in Bayern		
					Gesamt richtig pos. falsch pos.		
P1-3	1.5%	50	80	0.037	552	276	7183
P4-5			90	0.071			3611
			95	0.132			1815
		70	80	0.051	552	386	7182
			90	0.096			3634
			95	0.176			1807
		90	80	0.064	552	497	7268
			90	0.121			3610
			95	0.215			1814

Für die Röntgenuntersuchung der Lunge und die Mammographie wurde von weniger extremen Annahmen ausgegangen, unter anderem wegen der einfacheren Wiederholbarkeit dieser Untersuchungen. Die kurzfristige Wiederholung technisch unbefriedigender Bilder ist damit implizit als Teil des Tests inkorporiert. Wegen der im Vergleich

zum Skelett geringeren Prävalenz von Ereignissen ergibt sich auch unter optimistischeren Annahmen eine Relation von etwa 1 zu 3,5 für richtig vs. falsch-positive Befunde.

Während die Häufigkeit von Lungenbefunden durch Verbreiterung des Intervalls erhöht werden könnte, wäre für die Prävalenz bei der Mammographie eher eine Korrektur nach unten zu diskutieren, da die Häufigkeit entsprechender Befunde aufgrund der Zusammensetzung des analysierten Datenmaterials vermutlich überschätzt wurde (vgl. Abschn. 6.3.1).

Tab. 6.24: Ergebnisse der Gegenüberstellung von Metastasierungsmuster und Nachsorgestrategie: Mamma
Dargestellt sind der prädiktive Wert des positiven Tests und die erwarteten Befundhäufigkeiten in der bayerischen Bevölkerung

Programm	Prävalenz pro Test	Sens.	Spez.	Präd.Wert d.pos.Tests	Befunde pro Jahr in Bayern		
					Gesamt	richtig pos.	falsch pos.
P1-3	1,6%	50	80	0.039	302	151	3721
P4-5			90	0.075			1862
			95	0.140			928
		70	80	0.054	302	211	3696
			90	0.102			1858
			95	0.185			930
		90	80	0.068	302	272	3728
			90	0.128			1853
			95	0.226			932

Kritisch ist insbesondere die Nachsorge der Leber, deren Befundhäufigkeit den niedrigsten Wert aufweist. Innerhalb der ersten drei Jahre sind die halbjährliche (P1-3) und die jährliche (P'1-3) Untersuchung alternativ zu diskutieren. Für die Spezifität der sonographischen Leberuntersuchung wurde erneut ein breites Spektrum möglicher Werte angenommen, da entsprechende Hinweise in der Literatur vorliegen, wie oben zitiert wurde. Unter der optimistischsten Annahme ergibt sich ein Verhältnis von 1 zu 10 für richtig- und falsch-positive Befunde.

Tab. 6.25: Ergebnisse der Gegenüberstellung von Metastasierungsmuster und Nachsorgestrategie: Leber
Dargestellt sind der prädiktive Wert des positiven Tests und die erwarteten Befundhäufigkeiten in der bayerischen Bevölkerung

Programm	Prävalenz pro Test	Sens.	Spez.	Präd.Wert d.pos.Tests	Befunde pro Jahr in Bayern		
					Gesamt	richtig pos.	falsch pos.
Pl-3	0.0057	50	50	0.006	141	71	11762
			70	0.009			7818
			90	0.027			2559
		70	50	0.008	141	99	12276
			70	0.013			7516
			90	0.038			2506
		90	50	0.010	141	127	12573
			70	0.017			7343
			90	0.048			2518
	0.0113	50	50	0.011	142	71	6384
			70	0.019			3666
			90	0.054			1244
		70	50	0.016	142	99	6089
			70	0.026			3708
			90	0.074			1239
		90	50	0.020	142	127	6223
			70	0.033			3721
			90	0.093			1239
P4-5	0.0113	50	50	0.011	75	38	3417
			70	0.019			1962
			90	0.054			666
		70	50	0.016	75	53	3260
			70	0.026			1985
			90	0.074			663
		90	50	0.020	75	68	3332
			70	0.033			1993
			90	0.093			663

Alternativ zur Nachsorgestrategie des TZM wurde aufgrund des ungünstigsten prädiktiven Werts der Leber-sonographie eine selektive, befundgesteuerte Untersuchung der Leber simuliert.

Tab. 6.26: Alternative zur periodischen Beobachtung der Leber: Untersuchung nur bei positivem Skelett- und/oder Lungenbefund. Dargestellt sind der prädiktive Wert des positiven Tests und die erwarteten Befundhäufigkeiten in der bayerischen Bevölkerung

				Befunde pro Jahr in Bayern					
Pro- gramm pro Test	Prävalenz	Sens.	Spez.	Präd.Wert d.pos.Tests bei		Gesamt mit Strategie richtig Q entdeckbar* positiv* falsch positiv*			
				entw.Lunge oder Skelett	Lunge und Skelett				
Q1-3	50	80		0.217	0.428	142	75+29	38+15	137+20
0.1,		90		0.357	0.599				68+10
falls		95		0.526	0.749				34+ 5
entw.									
Lunge	70	80		0.280	0.511	142	75+29	53+20	136+19
oder		90		0.438	0.676				68+10
Skelett;		95		0.609	0.807				34+ 5
0.23,									
falls	90	80		0.333	0.573	142	75+29	68+26	136+19
beide		90		0.500	0.729				68+10
Organe		95		0.667	0.843				34+ 5
befallen									
Q3-4	50	80		0.217	0.428	75	40+16	20+ 8	72+11
		90		0.357	0.599				36+ 5
		95		0.526	0.749				18+ 3
	70	80		0.280	0.511	75	40+16	28+11	72+11
		90		0.438	0.676				36+ 5
		95		0.609	0.807				18+ 3
	90	80		0.333	0.573	75	40+16	36+14	72+10
		90		0.500	0.729				36+ 5
		95		0.667	0.843				18+ 3

* Linker Wert: entweder Skelett oder Lunge
rechter Wert: gleichzeitig Skelett und Lunge

Der Anteil von Lebermetastasen, die synchron bei Befall eines oder mehrerer anderer Organe manifest werden, ist weiterhin in einem hohen Prozentsatz erkennbar. Der Anteil falsch-positiver Befunde reduziert sich dagegen auf ein Minimum. Die günstigeren Annahmen über die Spezifität entsprechen der Annahme, daß bei manifester Skelett- bzw. Lungenmetastasierung die Betreuung des Patienten meist an ein spezialisiertes Zentrum übergeht. An dem gemäß Tabelle 6.26 erreichbaren Ergebnis sind die in Tabelle 6.25 ausgewiesenen Häufigkeiten zu messen. Es wird erkennbar, daß die hohe Rate falsch positiver Befunde überwiegend durch den Versuch entsteht, solitäre Lebermanifestationen frühzeitig zu entdecken.

Tab. 6.27: Erforderlicher Aufwand zur frühzeitigen Erfassung isolierten Leberbefalls: Gegenüberstellung der aus den Strategien P und Q resultierenden Befundhäufigkeit, jeweils für den Fall einer 90%igen Spezifität

Strategie	Sens.	Spez.	mögl.Bef.	Ergebnisse der regulären		Ergebnisse	
				Nachsorge (Strat. P bzw. P')		der Alternative Q	
				richtig-pos.	falsch-pos.	richtig pos.	falsch-pos.
P1-3	50	90	142	71	2559	53	78
	70	90	142	99	2506	73	78
	90	90	142	128	2518	94	78
P'1-3	50	90	142	71	1244	53	78
	70	90	142	99	1239	73	78
	90	90	142	128	1239	94	78
P4-5	50	90	75	38	666	28	41
	70	90	75	53	663	39	41
	90	90	75	68	663	50	41

Ausgehend von Strategie Q bedeutet der Übergang auf Strategie P bzw. P', daß einem Zugewinn von 35% richtig-positiver Befunde ein Anwachsen der Rate falsch-positiver Befunde auf das 32fache bzw. 16fache gegenübersteht.

6.3.4 Diskussion der Schätzung von Metastasierungswahrscheinlichkeiten

6.3.4.1 Probleme der Datenlage

Im vorangegangenen Abschnitt wurde versucht, im wesentlichen anhand der Daten eines Tumorverlaufsregisters die Situation der Tumornachsorge des Mammakarzinoms zu beschreiben und mit einer empfohlenen Nachsorgestrategie zu vergleichen. Die durchgeführten Schätzungen sollten in erster Linie demonstrieren, welche Qualität von Aussagen unter der derzeitigen Datenlage gewonnen werden kann. Die Diskussion prinzipieller, durch Rechentechniken nur bedingt behebbarer Limitierungen aufgrund der Zusammensetzung des Materials ist Gegenstand dieses Abschnitts.

Als einfachste Fehlerquelle im Datenmaterial ist die zentrumsinterne Selektion zu diskutieren. Hierunter werden Mechanismen verstanden, die bewirken, daß nicht alle im Zentrum behandelten Patienten durch die Dokumentation erfaßt werden. Wie im Abschnitt 6.3.1 ausgeführt wurde, scheinen rechnerisch erkennbare Selektionseffekte im prospektiven Kollektiv nicht vorzuliegen.

Nicht alle denkbaren Formen von zentrumsinterner Selektion manifestieren sich jedoch in rechnerisch erfaßbaren Verschiebungen innerhalb der Daten. Dies betrifft insbesondere stabile, d.h. zeitlich invariant wirksame Fehlerquellen, wie z.B. die Nichtteilnahme einer Abteilung eines Hauses mit systematisch abweichendem Patientengut. Wesentlicher als die nachträgliche Analyse - obwohl schwerer kommunizierbar - ist daher die adäquate Organisation der Erhebung, die systematische Selektion ausschließt. Bei Verwendung publizierter Daten kann in dieser Frage nur von Vermutungen ausgegangen werden. Auch wird das Problem durch Randomisierung nicht beseitigt. Ein wirksamer Schutz vor zentrumsinterner Selektion dürfte in der engen Kopplung von Dokumentation und Versorgung liegen. Ein Beispiel ist die Verwendung von Nachsorgepässen als organisatorische Basis der Nachsorge und zugleich zur Identifikation der Patientendaten. Dieses System schließt, wenn es erfolgreich arbeitet, bereits bei der Ersterfassung Selektion in der Dokumentation nahezu aus (26).

Auch bei zentrumintern lückenloser Erfassung ist die Repräsentativität der im Zentrum versorgten Patienten nicht gewährleistet. Multizentrische Ansätze können der Repräsentativität näherkommen als monozentrische. Zur Überprüfung bietet sich der interinstitutionelle Vergleich von Kriterien an. Dies wurde im Rahmen der gegebenen Möglichkeiten wahrgenommen. Unter der weiteren Annahme, daß prognostische Faktoren und soziodemographische Merkmale korreliert sind, könnten durch den Vergleich von Basisdaten der Stichprobe mit den Parametern der Grundgesamtheit indirekte Schlüsse gezogen werden. Dies wurde durch Vergleich mit dem Altersaufbau der im Saarland an malignen Tumoren der Mamma erkrankten Frauen versucht. Direkte Vergleichsdaten für Bayern selbst liegen bekanntlich nicht vor. Auch eine Überprüfung der Repräsentativität anhand der relevanten Merkmale ist undurchführbar, da weder für die Stichprobe noch für die Grundgesamtheit entsprechende Daten verfügbar sind. Letztlich können daher nur Stichproben geeigneter Größe und Struktur mit hinreichender Sicherheit als repräsentativ angesehen werden. Für das Krankengut eines Tumorzentrums allein ist diese Sicherheit nicht gegeben. Große versorgungsorientierte Register an Tumorzentren können jedoch zur Klärung der prognostischen Relevanz verschiedener Basismerkmale beitragen und dadurch Anforderungen an "Repräsentativität" präzisieren helfen.

Weitere Probleme stellen sich aufgrund der Unvollständigkeit der Verlaufbeobachtung. Zwar stehen mit der Sterbetafeltechnik Methoden zur Analyse unvollständig beobachteter Zeitdauerdaten zur Verfügung. Diese Techniken setzen jedoch die Unabhängigkeit von Zensierung und Zielereignis voraus, wie sie vorwiegend beim Erreichen des Studienendes ohne manifestes Zielereignis bei limitierter Beobachtungsdauer gegeben ist. Zensierung durch sonstigen Verlust von Patienten aus dem Follow-up ist nicht zwingend ein zufälliger Prozeß, im Gegenteil.

Für Kliniken, die selbst keine systematische Nachsorge für alle Patienten durchführen, ist im Verlauf mit einer Überschätzung der Ereignisrate zu rechnen. Der Zensierungsmechanismus ist nicht zufällig, da ereignisfreie Patienten in der Betreuung des Hausarztes verbleiben, progrediente oder fragliche Fälle dagegen der Klinik zugewiesen werden. Dieser Zensierungsmechanismus ist für das prospektive Kollektiv des Tumorregisters teilweise anzunehmen.

Für das Patientengut einer prospektiven Studie können umgekehrte Verhältnisse vorliegen. Wenn eine Patientin dazu angehalten wurde, über Jahre hinweg alle 3 Monate in der Klinik zur Nachsorge zu erscheinen, wie im Rahmen der Mailänder Studie, so könnte für die nach Ende der Studie auftretenden Verluste aus dem Follow-up mit größerer Wahrscheinlichkeit eine Tumorprogression verantwortlich sein, die die Patientin am weiteren Erscheinen hindert, denn für das Verhalten des Patienten ist der Unterschied zwischen Studie und Post-study Follow-up irrelevant. Deshalb wurden die ebenfalls publizierten 10-Jahres-Ergebnisse des Mailänder Kollektivs nicht berücksichtigt.

Vor einer etwaigen Extrapolation auf die Verhältnisse in einem Nachsorgeprogramm ist weiterhin zu bedenken, daß sich die Progressionsrate zwischen den Teilnehmern und den Nichtteilnehmern am Nachsorgeprogramm unterscheiden kann. Ursachen wurden in Abschnitt 3.5 genannt. Deshalb ist auch eine nach bestem Wissen und Ermessen aus zensierten Lebensdauerdaten abgeleitete Schätzung keine valide Vorhersage für die tatsächliche Ereignishäufigkeit bei den Teilnehmern an einem Nachsorgeprogramm.

Abschließend sind Fragen der Beobachtungsgleichheit zu erörtern, insbesondere im Hinblick auf die Übertragbarkeit der früher beobachteten Metastasierungsmuster auf zukünftige Verläufe. Bei der Auswahl der verwendeten Registerdaten bleiben verschiedentlich Verstöße gegen das Gebot der Beobachtungsgleichheit unbeobachtet. Dies schlägt sich insbesondere in einer Unschärfe der zeitlichen Zuordnung von Ereignissen nieder.

Typisch ist die Unschärfe der Datierung lang zurückliegender Ereignisse im retrospektiven Kollektiv. Hinzu kommt die Tendenz zur zeitlichen Identifizierung zurückliegender, im Abstand von Wochen oder Monaten aufgetretener Ereignisse ("1975 Lunge und Skelett"). Dadurch wird die Glaubwürdigkeit für die aus den Daten ermittelten Wahrscheinlichkeiten des gemeinsamen Auftretens verschiedener Metastasen relativiert. Als dritte Beeinträchtigung der Beobachtungsgleichheit in der Gegenüberstellung von prospektiv und retrospektiv dokumentierten Ereignissen ist die Kenntnis des weiteren Krankheitsverlaufs bei letzteren zu nennen. Die Interpretation von Befunden wird verändert, wenn Patienten noch Jahre nach

angeblicher Metastasierung frei von weiterer Progredienz sind. Von dieser Unschärfe ist im übrigen nicht nur das retrospektive Kollektiv betroffen, da auch Teile des prospektiven Kollektivs retrolektiv erfaßt sind.

Hinzu kommt die Abhängigkeit von Ereignis und benutztem Test, die in Abschnitt 3.3.5 erläutert wurde. Weder zwischen den Kollektiven noch innerhalb des prospektiven oder retrospektiven Kollektivs kann von der Vergleichbarkeit bzw. Einheitlichkeit der angewandten Diagnostikstrategien ausgegangen werden. Die Patienten wurden weder in einheitlichem Zeitraster untersucht noch wurden für alle Patienten vergleichbare Tests benutzt. Allein durch die Einführung von Computertomographie und Sonographie in den letzten Jahren wurden Zäsuren geschaffen, die an sich nicht retrospektiv übersprungen werden sollten. Welche Unterschiede zwischen der Lebermetastase von 1980 und der Lebermetastase von 1970 bestehen, kann aus den vorliegenden Daten nicht beantwortet werden. Wenn man den erheblich höheren Anteil des Leberbefalls im späteren Verlauf der Metastasierung oder die hohe Frequenz von Lebermetastasen in Obduktionsstatistiken berücksichtigt, so sind die beobachteten Häufigkeiten der Lebermetastasierung im retrospektiven Kollektiv auch mit der Annahme verträglich, daß Lebermetastasierung bei langsamem Größenwachstum der Metastasen verhältnismäßig früh im Verlauf beginnt und damit ähnliches Verhalten zeigt wie Skelettmetastasen. Unter dieser Annahme würden in den kommenden Jahren aufgrund der veränderten Diagnostik andere Verlaufsmuster entstehen, insbesondere wäre dann eine frühe Leberbeteiligung in einem höheren Prozentsatz zu erwarten, als in den verfügbaren Daten dokumentiert.

6.3.4.2 Alternativen zum gewählten Modell

Die mathematische Modellierung des direkt kaum beobachtbaren Prozesses des Tumorwachstums sowie der Wechselwirkungen zwischen Tumorwachstum und Früherkennung wird mit unterschiedlichen Akzenten in zahlreichen theoretischen Arbeiten versucht. Verfahren zur Modellierung der durch Screening erzeugten Effekte sind z.B. in (55) zusammengestellt. Wegen des Mangels an verwertbaren Daten ist die praktische Anwendbarkeit dieser Modelle limitiert, gelegentlich werden Ansätze zur Überwindung des Datenmangels disku-

tiert (z.B. 50). Teilweise sind diese Modelle auch auf die Tumornachsorge anwendbar, vereinzelt werden auch Probleme der Tumornachsorge gezielt aufgegriffen (19).

Das bislang umfassendste Modell wurde von EDDY (20) vorgelegt. Hier werden auch konkrete Wege aufgezeigt, wie die benötigten Parameter aus existierenden Beobachtungen approximiert werden können. Die Arbeitsweise dieses Modells wird zudem durch konkrete Anwendungen aus der Krebsfrüherkennung belegt. Wegen seiner verhältnismäßigen Praxisnähe und zugleich Allgemeinheit wäre EDDY's Modell vergleichsweise gut zur Beschreibung bzw. Simulation von Nachsorgeprogrammen geeignet.

Für die gegebene Anwendung führt jedoch der Abgleich von Modellanforderungen und Datenlage zum negativen Ergebnis. Zur Simulation der erwarteten Befundhäufigkeiten würde neben den Testcharakteristiken und der Inzidenz von Verlaufereignissen insbesondere die Dichte-Funktion des okkulten Intervalls des Tests, das ist der Zeitraum von der Entdeckbarkeit einer Metastase durch den Test bis zur klinischen Manifestation anhand von Zeichen oder Beschwerden, benötigt. Eine unmittelbare Beobachtung dieser Verteilung ist naturgemäß nicht möglich, weder beim Screening noch in der Nachsorge. EDDY empfiehlt daher, stattdessen die "Schatten" zu beobachten, die diese gesuchte Größe unter verschiedenen Blickwinkeln, d.h. unterschiedlichen Untersuchungsstrategien auf die beobachtbare Wirklichkeit wirft. Je mehr alternative Befundsequenzen aus unterschiedlichen Screeningprozeduren verfügbar sind, um so besser kann die gesuchte Verteilung demzufolge approximiert werden.

Voraussetzung zur Anwendung dieser Approximationstechniken ist also die Information, unter welchen Umweltbedingungen, insbesondere mit welchem Test beobachtete Ereignisse diagnostiziert wurden. An dieser Stelle versagen die verfügbaren Daten ihren Dienst. Bereits in der Diskussion des Materials wurde darauf hingewiesen, daß schon die zeitliche Eingrenzung des Auftretens von Befunden retrospektiven Unschärfen ausgesetzt ist. Informationen über die Art der Befundsicherung sind in noch geringerem Umfang verfügbar. Selbst indirekte Rückschlüsse, z.B. über die Kenntnis der üblicherweise in einer Klinikambulanz praktizierten Nachsorge

usw. sind nicht möglich, da die retrospektive Verlaufserfassung den Ort der erstmaligen Diagnosestellung nicht erkennen läßt.

Die Anwendung des EDDY-Modells hätte daher arbiträre Annahmen über die Verteilung des fraglichen Zeitintervalls erfordert. Die vermeintlich höhere Präzision des exakteren Modells hätte mit Vergrößerung der Unschärfe in den Eingangsdaten erkaufte werden müssen. Daher wurde auf das beschriebene, einfachere Modell zurückgegriffen.

Auch dieser Ansatz kommt nicht ohne Annahmen aus, die in Abschnitt 6.3.3.2 beschrieben wurden. Die entscheidende Voraussetzung lautet, daß alle durch Nachsorge zum Zeitpunkt t_i entdeckten Metastasen sonst bis zum Zeitpunkt t_{i+1} aufgetreten wären. Dies impliziert, daß durch die Nachsorge keine intervallübergreifenden Veränderungen der Modellkurve ausgelöst werden, so daß die Nachsorge im Intervall $[t_i, t_{i+1})$ nicht durch die Ergebnisse des vorhergehenden Intervalls $[t_{i-1}, t_i)$ beeinflusst wird. Der Effekt dieser Annahme ist eine mögliche Unterschätzung der Befundhäufigkeiten am Anfang der Nachsorgeperiode zugunsten der Überschätzung der später auftretenden Ereignisfrequenzen.

Welcher Anteil der im Folgeintervall erwarteten Befunde zum Zeitpunkt t_i effektiv entdeckbar ist, hängt vom Anteil bereits durch Früherkennung in der Nachsorge diagnostizierten Ereignisse in den benutzten Daten ab. Ist dieser hoch, so wird der Zeitverlauf auch durch Anwendung empfindlicher Tests nicht wesentlich nach links verlagerbar sein. In die Sensitivität des Tests fließt also neben der Trennschärfe, mit der erkennbare Befunde effektiv erkannt werden, auch die Struktur der Zielpopulation, insbesondere die Verteilung der Tumorgroße in den Zielorganen, ein (vgl. Abschnitt 3.3.2.1). Daher wurde das Intervall der Sensitivität nach unten mit dem Wert 50% eröffnet.

6.3.4.3 Aussagekraft der Modellrechnung und mögliche Folgerungen

Durch die erstellten Modellkurven zum Metastasierungsverlauf sollte die verfügbare Information genutzt werden. Zugleich beleuchtet die Analyse der Daten Probleme der Anlaufphase eines Tu-

morregisters, die sich über einen Zeitraum von mindestens 8 Jahren zu erstrecken scheint. Durch die Kombination von prospektiv und retrospektiv dokumentierten Verläufen wurden einige der angesprochenen Probleme verschärft, insbesondere die Frage der Vergleichbarkeit von Zielereignissen.

Andererseits ist auch in Zukunft damit zu rechnen, daß sich die diagnostischen Möglichkeiten zur Erfassung klinisch latenter Tumorprogressionen weiter verbessern werden. Deshalb werden auch in Zukunft Zweifel an der prospektiven Gültigkeit akutell ermittelter Verlaufskurven nicht auszuschließen sein.

Grenzen der erreichbaren Genauigkeit wurden sichtbar, nicht nur in der Diskussion der Modellannahmen. Auch die Festlegung des Parameters zur Beschreibung der Metastasierungswahrscheinlichkeit läßt Fragen offen. Ist das Modell der Exponentialverteilung zur Beschreibung der Metastasierung adäquat? Warum wurde $\lambda = 0.00025$ gewählt, wäre nicht 0.00026 vorzuziehen? Welche Veränderungen würden durch Integration der selteneren Metastasen wie ZNS oder Nebenniere bewirkt? Restlos befriedigende Begründungen können nicht gegeben werden, außer daß versucht wurde, die verfügbaren Beobachtungen "plausibel" zu interpretieren.

Dieses Vorgehen hat seinen Preis: die Interpretation der Ergebnisse kann und darf nicht blind gegen die Entstehung der Daten sein. Ableitbar sind vorwiegend Tendenzen, Relationen und Größenordnungen. Detailvorhersagen sind nicht gerechtfertigt. Jedoch soll trotz der dargestellten Limitierungen der Versuch einer Bewertung der Nachsorgeempfehlungen des TZM unternommen werden.

Klinik/Labor: Diese Untersuchungen wurden in der vorliegenden Auswertung nur indirekt berücksichtigt, als Abfangmechanismus für die im jeweiligen Test nicht diagnostizierten Ereignisse. Die Beurteilung von Sensitivität, Spezifität usw. klinischer Untersuchungen mußte allein schon deswegen unterbleiben, weil für das lokoregionäre Rezidiv bzw. die narbennahe Hautmetastase keine sinnvolle Verlaufskurve abgeleitet werden konnte (vgl. 6.2.4).

Andererseits wäre es verfehlt, die Bedeutung der aufmerksamen Beobachtung des Patienten auf eine oder zwei methodische Kenngrößen zu reduzieren. Die Wechselwirkung zwischen Klinik und Tests wurde in Abschnitt 3.3 diskutiert. Aus planerischer und methodischer Sicht bietet der regelmäßige Kontakt zum Patienten zusätzlich die Gewähr, daß rasch progrediente Befunde zeitgerecht erfaßt werden. Aus diesen Gründen erscheinen Änderungen an dieser - in allen Programmen nahezu einheitlichen - Komponente der Empfehlungen nicht angezeigt.

Skelett: Innerhalb der gewählten Bandbreite von 50 - 90% für die Spezifität der Skelettszintigraphie können aus Tabelle 6.22 unterschiedliche Schlußfolgerungen gezogen werden. Entscheidend für die Bewertung sind Annahmen über die tatsächlich erreichbare Spezifität der Untersuchung unter den Bedingungen der Nachsorge. Hier sprechen einige Gründe für optimistischere Prognosen, z.B. die Möglichkeit zum Vergleich aktueller Befunde mit älteren Bildern.

Etwas im Gegensatz zu den Annahmen, unter denen die Modellrechnung durchgeführt wurde, müßte dann die Wahrscheinlichkeit des Auftretens bzw. der Diagnostizierung einer Skelettmetastase im ersten Verlaufsabschnitt geringer sein und später rascher ansteigen. Ein entsprechendes Metastasierungsmuster wurde in (40) in der Tat beobachtet.

Eine defensive, d.h. hochspezifische Interpretation der Skelettszintigraphie wäre demnach von Anfang an zu erzielen, wobei die Sensitivität von anfänglich ungünstigeren auf allmählich gute Werte ansteigen könnte. Die Prävalenz erkennbarer Skelettmetastasen könnte sich damit, abweichend vom Modell der Tabelle 6.23, in Richtung des mittleren bis späteren Verlaufsabschnitts verlagern. Damit könnte die periodische Skelettszintigraphie über den gesamten Verlauf der Nachsorge mit niedriger Falsch-Positiv-Rate arbeiten und bei anfänglich geringerer Ausbeute zunehmend an Sensitivität gewinnen.

Lunge/Pleura: Die Röntgenuntersuchung des Thorax ist die einfachste und zugleich häufigste Untersuchung in der Nachsorgestrategie des Tumorzentrums. Eine Begründung, weshalb die

Skelettszintigraphie im zweiten Abschnitt der Nachsorge (P3-4) in der Frequenz reduziert wird, während die 6monatige Thoraxkontrolle bis zum Ende beibehalten wird, ist aus den Daten jedoch nicht unmittelbar ersichtlich.

Kontralaterale Mamma: Trotz der mutmaßlichen Überschätzung der Frequenz von Tumormanifestationen in der kontralateralen Mamma ist die ermittelte Häufigkeit dieses Ereignisses gering. Die Diskussion über die Zweckmäßigkeit dieser Strategiekomponente anhand der Daten aus Tabelle 6.24 wird unmittelbar durch die Frage nach dem Ziel der Untersuchung beeinflusst.

Die Mammographie zur frühzeitigen Entdeckung von Mammametastasen steht in Konkurrenz zur klinischen Untersuchung, die ebenfalls zur Entdeckung von Mammatumoren führen kann. Mammographie nur bei auffälligem Tastbefund wäre eine denkbare Alternative zum vorgeschlagenen Vorgehen. Der höhere Aufwand einer jährlichen Mammographie auch bei klinisch unauffälliger Mamma muß letztlich im Hinblick auf den erwarteten medizinischen Nutzen diskutiert werden.

Verzichtete man auf die jährliche Mammographie, so wäre für die Frage einer etwaigen Tumorprogression im Zeitraum zwischen der möglichen mammographischen Entdeckbarkeit der Mammametastase und der tatsächlichen Entdeckung durch klinische Untersuchung zunächst auf die Häufigkeiten in Tabelle 6.19 zu verweisen. Hier weisen 94% der Patientinnen bei positivem Mammabefund keine weiteren Fernmetastasen auf. Diese Zahl kann auf zwei verschiedene Arten interpretiert werden. Nimmt man an, daß alle diese Befunde frühzeitig durch Mammographie gesichert wurden, so wäre die Untersuchung der kontralateralen Brust als ideale Maßnahme zur frühzeitigen Feststellung einer Tumorprogression anzusehen. Es müßte dann angenommen werden, daß zwischen der früheren Entdeckung durch Mammographie und der Abwesenheit weiterer Tumormetastasen ein Kausalzusammenhang bestehen könnte und daß bei Verschleppung der Diagnose der Mammametastase zwischenzeitlich weitere Tumormetastasen auftreten könnten.

Träfe diese Interpretation zu, müßten allerdings auch zahlreiche Mammametastasen beobachtbar sein, wenn Skelett-, Lungen- oder Lebermetastasen dignostiziert werden. Dies ist jedoch aus den Daten der Tabelle 6.19 nicht ersichtlich.

Eine andere Interpretation ist daher wahrscheinlicher, nämlich daß der Zeitverlust zwischen potentieller Entdeckung durch Mammographie und tatsächlicher Entdeckung durch klinische Untersuchung kaum zur Ausweitung der Tumorprogression an anderen Orten beiträgt. Durch welche physiologischen Modelle dies auch erklärbar sein mag, eine jährliche Mammographie zur Metastasensuche erscheint aufgrund der vorliegenden Daten nicht zwingend geboten.

Neben der Früherkennung einer Tumorprogression wird die frühzeitige Entdeckung von Zweitmalignomen beabsichtigt. In diesem Kontext ist die Mammographie als Screeningmaßnahme an einem High-Risk-Kollektiv zu betrachten. Dieses Screening sollte dann allerdings nicht auf den 5-Jahreszeitraum nach Diagnosestellung des ersten Primärtumors beschränkt bleiben, da das erhöhte Zweitmalignomrisiko gegebenenfalls lebenslang fortbesteht.

Leber: Angesichts der in Tabelle 6.25/6.26 vorliegenden Werte und der nach (54) gerade für Lebermetastasen eines Mammakarzinoms unbefriedigenden Sensitivität der Oberbauchsonographie scheint die zurückhaltendere Strategie naheliegend zu sein. Offene Fragen bleiben allerdings zur Validität der Daten - aus heutiger Sicht - zurück, denn ein großer Teil dieser Verläufe wurde vor dem Beginn und in der Frühphase der Sonographie- und Computertomographie-Ära beobachtet. In welcher Frequenz die Leberszintigraphie durchgeführt wurde, ist retrospektiv nicht eruierbar. Klinisch können Lebermetastasen eines Mammakarzinoms längere Zeit stumm bleiben. Dieser Einwand wird durch den Befund unterstützt, daß synchron mit der Diagnose einer Skelettmetastase Lungen- und Leberbefunde in vergleichbarer Frequenz (19% versus 12%) auftraten, daß also bei der durch den Skelettbefund ausgelösten intensiveren Tumorsuche Lebermetastasen in nicht unbeträchtlicher Frequenz zu Tage traten. Damit allerdings dürfte der größte Teil der tatsächlich vorhandenen Lebermetastasen entdeckt sein, so daß in den Daten des Tumorregisters weniger das Auftreten von Lebermetastasen insgesamt, sondern eher deren alleiniges Auftreten als erste Manife-

station unterschätzt wäre. Deshalb bleibt auch in Würdigung dieser Einwände Veranlassung zur Zurückhaltung und Erwägung einer Alternativstrategie, wie sie in Tabelle 6.26 simuliert wurde.

Gesamtbeurteilung:

Zusammenfassend ergeben sich Gesichtspunkte, durch die die Nachsorgestrategie des TZM zum Mammakarzinom gestützt wird. Einwände sind im wesentlichen bezüglich zwei der vier untersuchten Strategiekomponenten zu diskutieren. Aufgrund der durchgeführten Modellrechnungen ist nicht auszuschließen, daß durch einzelne Programmteile ein zu aufwendiger Standard definiert wurde.

Eine abschließende Beurteilung der Empfehlungen aufgrund der durchgeführten Modellrechnungen allein ist allerdings nicht ableitbar. Dies konnte, gemessen an den Forderungen der Abschnitte 3, 4 und 5, auch nicht erwartet werden.

Der praktische Nutzen derartiger Simulationen betrifft insbesondere zwei Bereiche. Erstens dürften multizentrische, verlaufsorientiert arbeitende Register wie das TRM, ungeachtet der diskutierten Limitierungen, mit zu den ergiebigsten Datenquellen über Krankheitsverläufe zählen. Darauf aufsetzende Simulationen zeigen zumindest Größenordnungen auf und weisen auf mögliche Schwachstellen einer Nachsorgestrategie hin. Der zweite Effekt ist darin zu sehen, daß ein Erwartungshorizont abgesteckt wird, an dem die Trefferraten zukünftiger, großräumig praktizierter Nachsorgeprogramme gemessen werden können.

Diese Aspekte unterstreichen damit zugleich die Notwendigkeit der Durchführung prospektiver Studien, denn nur durch eine systematische, die Nachsorge begleitende und über das tumorfreie Intervall hinausweisende Dokumentation wird sich die Qualität der verfügbaren Informationen nachhaltig verbessern lassen.

7. Zusammenfassung

Der Stellenwert der Tumornachsorge für die klinische Onkologie wird durch die Vielfalt publizierter Empfehlungen belegt. Demgegenüber fand das Thema unter Biometrikern und Epidemiologen bislang wenig Beachtung. In der vorliegenden Arbeit wurde aus methodischer Sicht zu Fragen der Wirkung und Wirksamkeit diagnostischer Nachsorgeprogramme Stellung genommen. Hierzu wurden in einem theoretischen Teil (Abschn. 2-5) methodische Aspekte der Planung und Bewertung untersucht und in einer Anforderungsliste zusammengefaßt. Mit verschiedenen Modellrechnungen (Abschn. 6) wurden anschließend Planungsgrößen für die Durchführung von Nachsorgeprogrammen in Bayern quantitativ ermittelt.

Im 2. Abschnitt wurde eine Bestandsaufnahme zur Tumornachsorge vorgenommen. Unter Nachsorgeprogramm wurde hierbei eine vorgeschriebene Abfolge diagnostischer Tests verstanden, die für eine definierte Zielgruppe in einem definierten Umfeld anwendbar ist.

Im 3. Abschnitt stand die Planung bzw. ex-ante-Evaluation von Nachsorgeprogrammen im Mittelpunkt. Behandelt wurden Fragen der Art, Anzahl, Abfolge und Treffsicherheit der zu benutzenden diagnostischen Tests sowie der Häufigkeit der gesuchten Ereignisse.

Methoden zur Validierung diagnostischer Tests sowie Auswirkungen unterschiedlicher Teststrategien auf die Ergebnisse der Befundung wurden zuerst diskutiert. Anschließend wurden Faktoren zusammengestellt, durch die der von Sensitivität, Spezifität und Prävalenz abhängige prädiktive Wert diagnostischer Tests im Rahmen der Nachsorge beeinflußt werden kann.

Für Sensitivität und Spezifität der diagnostischen Tests ist in der Nachsorge mit ungünstigeren Werten zu rechnen, als auf Grund nachsorgeunabhängig ermittelter Güteindizes erwartet werden könnte. Durch geeignete Maßnahmen kann den Ursachen dieser Verschlechterung jedoch gegengesteuert und damit die Wirkung von Nachsorgeprogrammen verbessert werden. Strategien wurden vorgeschlagen.

Weiter wurden Einflußfaktoren für die Prävalenz der Zielereignisse aufgezeigt. Die Anforderungen an die Definition geeigneter Zielereignisse wurden präzisiert.

Im 4. Abschnitt wurden methodische Ansätze zur Abschätzung der Wirksamkeit, d.h. des Einflusses von Nachsorgemaßnahmen auf vorgegebene Zielkriterien, untersucht.

Als Zielkriterium bietet sich, neben subjektiven Parametern, insbesondere das Überleben ab erstmaliger Diagnosestellung an. Allerdings liegen Anfangs- und Endpunkt dieser Zeitspanne außerhalb des Nachsorgeintervalls. Daraus ergeben sich besondere Anforderungen an die Durchführung und Organisation von Studien.

Probleme für die Durchführung randomisierter Studien werden bezüglich der Auswahl der zu prüfenden Alternativen, der Akzeptanz der Studie sowie der Konstanz der Voraussetzungen gesehen. Ergänzend oder alternativ wurden daher Fall-Kontrollstudien in der Phase nach Einführung des Programms vorgeschlagen. Kriterien für die Durchführung solcher Studien wurden erarbeitet.

Im 5. Abschnitt wurde eine ausführliche Anforderungsliste (Tab. 5.2) vorgelegt, die die methodischen Einzelergebnisse des theoretischen Teils zusammenfaßt. Ausgehend vom 10-Punkte-Katalog der WHO (71) für Screenigprogramme werden für die Nachsorge substantielle Modifikationen erforderlich. Die praktische Erfüllbarkeit der entwickelten Anforderungsliste wurde durch Skizzierung einer Strategie zum Auf- und Ausbau von Nachsorgeprogrammen aufgezeigt.

Im 6. Abschnitt wurden in einer Folge von Modellrechnungen Planungsparameter quantitativ geschätzt.

In einer Extrapolation bis zum Jahr 2030 wurde die Entwicklung der Prävalenz Tumorkranker simuliert. Allein aufgrund der Verschiebungen im Altersaufbau ergibt sich eine prognostizierbare Steigerung der Inzidenz (Crude Rate) um 30 - 40%.

Die quantitative Beschreibung der aktuellen Nachsorgesituation ergab folgendes: Bei einer allgemeinen Nachsorge mit 14 Untersuchungen in 5 Jahren (4-3-3-2-2) ergäben sich rechnerisch für Bayern etwa 350.000 mögliche Arztkontakte jährlich. Reduziert um den geschätzten Anteil Nichtteilnehmer und/oder nicht tumorfreien Patienten wird mit jährlich etwa 170.000 effektiv durchzuführenden Nachsorgeuntersuchungen zu rechnen sein.

Zuletzt wurden am Beispiel Mammakarzinom Modellrechnungen zum natürlichen Krankheitsverlauf durchgeführt. Als praktisches Ergebnis ist festzuhalten, daß sämtliche Untersuchungen innerhalb eines realistischen Nachsorgeprogramms mit niedriger Prävalenz arbeiten, d.h. die erwartete Relation von falsch- versus richtig-positiven Befunden als ungünstig bezeichnet werden muß.

Simulationen helfen Fragestellungen zu präzisieren. Sie unterstreichen damit die Notwendigkeit prospektiver Untersuchungen. Nur durch eine systematische, die Nachsorge begleitende und über das tumorfreie Intervall hinausgehende Dokumentation kann die Qualität der verfügbaren Daten verbessert werden.

8. Literatur

- 1 Arnold, H.: Kosten und Nutzen der Krebsfrüherkennung. Öff. Gesundh.-Wesen 40, 1978, 329-338
- 2 Bayerisches Statistisches Landesamt: Statistische Berichte, A IV 3 - J/83. Selbstverlag des Bayerischen Statistischen Landesamtes, 1984
- 3 Bayerisches Statistisches Landesamt: Mortalitätsstatistik 1979-81 für Südbayern. Selbstverlag des Bayerischen Statistischen Landesamtes, 1982
- 4 Bayerisches Statistisches Landesamt: Altersstruktur der Bevölkerung Bayerns, Stand 31.12.81. Selbstverlag des Bayerischen Statistischen Landesamtes, 1982
- 5 Buser, K., Oeter, K.: Utopie Gesundheit. Münch. Med. Wschr. 123, 1981, 1843-1846
- 6 Camphausen, B.: Auswirkungen demographischer Prozesse auf die Berufe und die Kosten im Gesundheitswesen. Springer, Heidelberg, 1983
- 7 Charlson, M.E., Feinstein, A.R.: Rapid growth rate: a method of identifying node-negative breast cancer patients with a high risk of recurrence. J. Chron. Dis. 36, 1983, 847-853
- 8 Cochrane J.P., Williams, J.T., Faber, R.G., Slack, W.W.: Value of outpatient follow-up after curative surgery for carcinoma of the large bowel. Br. Med. J. 1 (6214), 1980, 593-595
- 9 Cohn, V.: Over 50% of Cancer Patients in U.S. said to survive 5 Years. Int. Herald Trib., 30.11.83
- 10 Collatz, J., Malzahn, P., Schmidt, E.: Erreichen die gesetzlichen Früherkennungsuntersuchungen für Säuglinge und Kleinkinder ihre Zielgruppen? Öff. Gesundh.-Wesen 41, 1979, 173-190
- 11 Collen, M.F.: Medical technology assessment: An emerging evaluation methodology. Meth. Inform. Med. 23, 1984, 1-3
- 12 Cox, D.R.: Regression models and lifetables. J. R. Stat. Soc. 34, Ser. B, 1972, 187-220
- 13 Dantzig, G.B.: Lineare Programmierung und Erweiterungen. Springer, Berlin-Heidelberg-New York, 1966
- 14 Day, E.: Is the periodic health examination worthwhile? Cancer 47, 1981, 1210-1214
- 15 De Vita V.T.: Principles of cancer therapy (in: Harrison's Principles of internal medicine, eighth edition), McGraw Hill, Kogakusha, 1977

- 16 Demling, L., Oeser, H., Schwartz, F.W., Soost, H.J.: Expertendiskussion: Welchen vernünftigen Zweifel kann es an der Krebsfrüherkennung geben? Münch. Med. Wschr. 126, 1984, 26-41
- 17 Derra, E., Elmendorf, H.: Spontanregression und Wachstumsverlangsamung von histologisch erwiesenen malignen Tumoren ohne oder ohne ausreichende kausale Therapie. Langenbecks Arch. Klin. Chir. 311, 1965, 89-99.
- 18 Dümmling, K.: 10 Jahre Computertomographie - Ein Rückblick. Electromedica 52, 1984, 13-28
- 19 Dwyer, A.J., Prewitt J.M.S, Ecker, J.G., Plunkett, J.: Use of the hazard rate to schedule follow-up exams efficiently. Med. Decis. Making 3, 1983, 229-244
- 20 Eddy, D.M.: Screening for cancer. Prentice-Hall, Englewood Cliffs, 1980
- 21 Eddy, D.M.: A mathematical model for timing repeated medical tests. Med. Decis. Making 3, 1983, 45-62
- 22 Fayers, P.M., Jones, D.R.: Measuring and analyzing quality of life in cancer clinical trials: A review. Stat. Med. 2, 1983, 429-446
- 23 Fletcher, R.H., Fletcher, S.W., Wagner, E.H.: Clinical epidemiology - the essentials. Williams & Wilkins, Baltimore-London, 1982
- 24 Galen, K.S., Gambino, S.R.: Beyond normality. The predictive value and efficiency of medical diagnosis. John Wiley & Sons, New York, 1975
- 25 Hölzel, D., Schubert-Fritschle, G., Thieme, Ch.: Tumornachsorge als Herausforderung an die Kooperationsfähigkeit in der Medizin. Jahrestagung der Dt. Ges. f. Soz-Med., München, 1982
- 26 Hölzel, D., Schubert-Fritschle, G., Thieme, Ch.: Klinikübergreifende Tumorverlaufsdokumentation. Springer, Heidelberg, 1984
- 27 Karnofsky, D.A. et al.: The use of nitrogen mustards in the palliative treatment of carcinoma. Cancer 1, 1948, 634-656
- 28 Kleinbaum, D.G., Muller, K.E.: Assessing errors from statistical decision making in epidemiologic studies. Manuskript, im Druck.
- 29 Klimm, H.D.: Tumornachsorge durch den Hausarzt. Z. Allg. Med. 60, 1984, 113-119
- 30 Koch, B.L., Voss, A.C., Ahlemann, L.M.: Prognose des Rezidivs beim operierten und nachbestrahlten Mammakarzinom. Strahlentherapie 156, 1980, 750-753

- 31 Köbberling, J., Windeler, J.: Der Test auf okkultes Blut im Stuhl. Studie zum Aussagewert für die Früherkennung kolorektaler Karzinome. Thieme Copythek, im Druck, 1985
- 32 Krebsfrüherkennung: Künftig gezielter vorgehen. Auszüge aus der BMFT-Studie zum Krebsfrüherkennungsprogramm. Münch. Med. Wschr. 126, 1984, 33-34
- 33 La Vecchia, C. et al.: Cervical neoplasia: Quantitative estimates from a case-control study. Lancet, 1984, 779-782
- 34 Lee, Y.T.N.: Bone scanning in patients with early breast carcinoma: Should it be a routine staging procedure? Cancer 47, 1981, 486-495
- 35 Leonhardt, A., Schuster, H.: Hinweise für die Führung der Dokumentationsbögen. Bay. Ärzteblatt, 1978, 489
- 36 Lieven, H. v.: Spezielle Strahlentherapie maligner Tumoren, Teil 2: Mamma-Tumoren (aus: Handbuch der medizinischen Radiologie Bd. XIX), Springer, Berlin-Heidelberg, 1982
- 37 Lock, W.: Epidemiologie der Krebserkrankung - 105 Jahre Mortalitätsstatistik. Münch. Med. Wschr. 123, 1981, 1491-1497
- 38 Maier, B.J.: On the lead time of a periodic screening program. Meth. Inform. Med. 22, 1983, 45-50
- 39 McNeil, B.J., Weichselbaum, R., Pauker, S.G.: Fallacy of the five-year survival in lung cancer. New Engl. J. Med. 299, 1978, 1397-1401
- 40 McNeil, B.J., Hanley, J.: Analysis of serial radionuclide bone images in osteosarcoma and breast carcinoma. Radiology 135, 1980, 171-176
- 41 McNeil, B.J., Weichselbaum, R., Pauker, S.G.: Speech and Survival. Tradeoffs between quality and quantity of life in laryngeal cancer. New Engl. J. Med. 305, 1981, 982-987
- 42 Moya, D.J.: The effectiveness of examinations in the early detection of cervical cancer. GMDS-Symposium, Freiburg, 10./11.9. 1983
- 43 Nagel, G.A.: Adjuvant Chemotherapy of breast cancer and related problems. J. Cancer Res. and Clin. Oncol. 107 (Supplement), 1984, 10
- 44 Neyman, J.: Outline of statistical treatment of the problem of diagnosis. Public Health Reports 62, 1947, 1449-1456
- 45 Ober, K.G.: Krebsregister - Möglichkeiten und Gefahren. Geburtsh. u. Frauenheilk. 43, 1983, 14-17
- 46 Oeser, H.: Krebs: Schicksal oder Verschulden? Thieme, Stuttgart, 1979

- 47 Politser, P.: Reliability, decision rules, and the value of repeated tests. *Med. Decis. Making* 2, 1982, 47-69
- 48 Prorok, P.C., Hankey, B.F., Bundy, B.N.: Concepts and problems in the evaluation of screening programs. *J. Chron. Dis.* 34, 1981, 159-171
- 49 Ransohoff, D.F., Feinstein, A.R.: Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New Engl. J. Med.* 299, 1978, 926-930
- 50 Robra, B.P., Schwartz, F.W.: Theoretical and practical criteria for the evaluation of screening programmes. *GMDS-Symposium, Freiburg*, 10./11.9. 1983
- 51 Sackett, D.L.: Clinical diagnosis and the clinical laboratory. *Clin. Investig. Med.* 1, 1978, 37-43
- 52 Sackett, D.L.: Bias in analytic research. *J. Chron. Dis.* 32, 1979, 51-63
- 53 Sauer, H. (Hrsg.): Empfehlungen zur Diagnostik, Therapie und Nachsorge: Mammakarzinom. Selbstverlag des Tumorzentrums München, 1983
- 54 Schölmerich, J., Volk, B.A., Neuner, C., Fröhlich, J., Gerok, W.: Aussagefähigkeit der Sonografie bei Lebermetastasen. *Dtsch. Med. Wschr.* 109, 1984, 326-329
- 55 Schubert, U., Seiffert, G.: Some aspects of mathematical screening models. *GMDS-Symposium, Freiburg*, 10./11.9. 1983
- 56 Seidman, H.: Statistical and epidemiological data on cancer of the breast. *American Cancer Soc.*, No. 3017, 1980
- 57 Shapiro, S.: Statistical evidence for mass screening for breast cancer and some remaining issues. *Cancer Detect. and Prev.* 1, 1976, 347-363
- 58 Siegrist, J., Bertram, H.: Schichtspezifische Variation des Krankheitsverhaltens. *Soziale Welt* 21/22, 1970/71, 206-218
- 59 Spitzer, W. et al.: Measuring the quality of life of cancer patients. *J. Chron. Dis.* 34, 1981, 585-597
- 60 Statistisches Amt des Saarlandes: Saarländische Krebsdokumentation 1979-1981. Selbstverlag des Statistischen Amtes des Saarlandes, 1983
- 61 Statistisches Bundesamt: Statistisches Jahrbuch 1981 für die Bundesrepublik Deutschland. Verlag W. Kohlhammer, Wiesbaden, 1981
- 62 Strender, L.E., Lagergren, C., Wallgren, S., Liljeval, S.: Role of bone scans in the initial assessment of operable patients with breast carcinoma. *Acta Radiol. Oncol.* 20, 1981, 187-191

- 63 Tallis, G.M., O'Neill, T.J.: Analysis of the detection process for breast cancer. *Biom. J.* 23, 1981, 203-217
- 64 UICC: TNM: Klassifikation maligner Tumoren, 3. Auflage. Springer, Berlin-Heidelberg-New York, 1979
- 65 Valagussa, P., Bonadonna, G., Veronesi, U.: Patterns of relapse and survival following radical mastectomy. *Cancer* 41, 1978, 1170-1178
- 66 Wagner, G., Becker, N.: Die Krebssterblichkeit in Mitteleuropa. *Dt. Ärzteblatt* 79, 1982, 37-44
- 67 Wagner, G., Grundmann, E. (Hrsg.): Basisdokumentation für Tumorkranke. Springer, Berlin-Heidelberg-New York, 1983
- 68 Walter, E.: Testing and estimating the sensitivity and the specificity of diagnostic procedures. *GMDS-Symposium, Freiburg*, 10./11.9. 1983.
- 69 Weber, F.: Evaluation of Cancer Screening. *GMDS-Symposium, Freiburg*, 10./11.9. 1983
- 70 Weibel, W.: 25 Jahre Wertheimscher Karzinomoperation. *Wiener Kl. Wschr.* 25, 1925, 716-719
- 71 Wilson, J.M.G., Jungner, G.: Principles and practice of screening for disease. *WHO: Public Health Papers* 34, 1968
- 72 Yates, J.W., Chalmer, B., McKegney P.: Evaluation of patients with advanced cancer using the Karnofsky performance status. *Cancer* 45, 1980, 2220-2224
- 73 Yerushalmy, J.: Statistical problems in assessing methods of medical diagnosis, with special reference to X-Ray techniques. *Public Health Reports* 62, 1947, 1432-1449
- 74 Zelen, M.: Theory of early detection of cancer in the general population. In: *Breast cancer: Trends in research and treatment.* Raven Press, New York, 1976

Mein Dank gilt Herrn Prof. Dr. K. Überla, unter dessen fachlicher Anleitung ich seit 1977 am ISB tätig bin. Seine Anregungen und Diskussionsbeiträge gaben dieser Arbeit wichtige Impulse.

Mein Dank gilt weiter Herrn Prof. Dr. D. Hölzel, der mir als geduldiger Vorgesetzter und engagierter Betreuer jederzeit beratend zur Seite stand.

Mein Dank gilt auch Frau Th. Goebel, Frau J. Hyde und Frau G. Kornely, die das Manuskript und die Abbildungen in die vorliegende Form gebracht haben.

LEBENS LAUF

Am 28.1.1952 wurde ich in Niederpöcking als Sohn des technischen Übersetzers Johannes Thieme und seiner Ehefrau Elisabeth, geb. Heinze, geboren.

1958 - 1962	Besuch der Volksschule in München
1962 - 1971	Besuch des humanistischen Ludwigs-Gymnasiums in München
Mai 1971	Abitur
WS 1971-72	Immatrikulation an der Technischen Universität München, Studienfach Informatik/Mathematik
SS 1973	Vordiplom
Mai 1977	Diplomprüfung

Seit November 1977 bin ich als wissenschaftlicher Angestellter am Institut für Medizinische Informationsverarbeitung, Statistik und Biomathematik der Ludwigs-Maximilians-Universität München beschäftigt.

Ich bin verheiratet und habe zwei Kinder.

